УДК 519.22
ББК 22.172
К 55

Кобзарь А. И. **Прикладная математическая статистика. Для инженеров и научных работников.** — М.: ФИЗМАТЛИТ, 2006. — 816 с. — ISBN 5-9221-0707-0.

В книге рассматриваются способы анализа наблюдений методами математической статистики. Последовательно на языке, доступном специалисту — не математику, излагаются современные методы анализа распределений вероятностей, оценки параметров распределений, проверки статистических гипотез, оценки связей между случайными величинами, планирования статистического эксперимента. Основное внимание уделено пояснению примеров применения методов современной математической статистики.

Книга предназначена для инженеров, исследователей, экономистов, медиков, аспирантов и студентов, желающих быстро, экономично и на высоком профессиональном уровне использовать весь арсенал современной математической статистики для решения своих прикладных задач.

ISBN 5-9221-0707-0

Посвящаю моей жене,
терпению которой
обязана эта книга

СОДЕРЖАНИЕ

Всякая вещь есть форма проявления беспредельного разнообразия.

Многие вещи нам непонятны не потому, что наши понятия слабы; но потому, что сии вещи не входят в круг наших понятий.

Козьма Прутков

О математической статистике и об этой книге

(обращение к читателю)

Что такое математическая статистика и зачем она нам?

Здравствуй, уважаемый читатель! Кто бы ты ни был — инженер, медик, экономист, агроном, биолог, психолог или географ, каждый день и каждый час ты имеешь дело с потоком данных, обрушивающихся на тебя. С их помощью окружающий нас мир пытается поведать о себе. Результатами испытаний прибора сообщить инженеру о том, что́ он создал; сведениями о заболеваниях рассказать медику о результатах его работы; информацией о работе промышленности заставить экономиста еще раз проверить эффективность экономической системы. Так или иначе, каждый из нас, оглядываясь в прошлое или заглядывая в будущее, не уйдет от необходимости получать информацию и извлекать из нее ответы на свои многочисленные вопросы.

Казалось бы, чего проще — взглянул инженер на результаты испытаний прибора и выявил все свои недоработки, медик получил результаты анализов и безошибочно поставил правильный диагноз. Однако горький опыт подсказывает, что это далеко не всегда так. Оказывается, что, наблюдая одно и то же явление, мы будем получать все время разные результаты. Это — проявление могущества Его Величества Случая. Слово *случай*, такое прозрачное для статистика-профессионала, остается для большинства инженеров промышленности, медиков, биологов и экономистов символом вмешательства темных, не поддающихся контролю сил. Отчасти это является интуитивной реакцией „здравого смысла“ на двойственность и взаимообусловленность понятийной пары „случайность — детерминированность“.

Любое событие (или цепь событий) наблюдатель справедливо считает проявлением реальности. С точки зрения теории вероятностей любое наблюдаемое событие уникально, неповторимо и, стало быть, формально невероятно, ибо для математической случайности нет нулевой и единичной вероятностей. Однако версия математической случайности предполагает, что наблюдаемая последовательность событий является частью более общей последовательности громадного периода, в которой наблюдаемая последовательность содержится много раз. Так детерминированность здравого смысла соприкасается со случайностью математической абстракции.

Возможности человека, слава Богу, будут всегда ограничены. Он не сможет повторить один и тот же опыт бесконечное число раз, а поэтому он никогда не узнает все обо всем, и ему всегда придется, принимая решение, исходить из своего нелегкого опыта. Однако надежды его не так уж беспочвенны, ибо еще гениальный В. Шекспир отметил, что „непременно за шалой случайностью, радуясь своей странноватой склонности, как свинья в грязи бескрайности, прикорнула наглейшая определенность“.

Как поймать за хвост эту „наглейшую определенность“, как увидеть что-то осмысленное в обрушивающейся на нас лавине информации, чем защититься от потока досаждающих случайностей?

Такой инструмент человек нашел более 300 лет тому назад: это математическая статистика — теория познания мира через опыт. Ее приемы и правила позволяют, располагая противоречивыми результатами наблюдений, выбрать из всех гипотез наиболее правдоподобную.

Один из основателей и корифеев математики случайностей Блез Паскаль определил ее как „учение, объединяющее точность математических доказательств с неопределенностью случая и примиряющее эти, казалось бы, противоречивые элементы“. Однако у неискушенного в премудростях науки современника статистика ассоциируется скорее с метким афоризмом Дизраэли — „есть ложь, большая ложь и статистика“, чем с изящным инструментом, крайне необходимым ему в работе. В Англии конца 18-го века описательная статистика получила знаменательное название „политическая арифметика“. Вот такой „политической арифметикой“, способной на потребу политикам „гармонизировать“ любые данные, и осталась для многих статистика. Подобно „доброму человеку из Сезуана“ Б. Брехта, мы привыкли к тому, что „дурной конец заранее отброшен, он должен, должен, должен быть хорошим“.

Ясно, что такие призывы никак не могут быть питательной средой математической статистики. Ее питательной средой и ресурсной базой являются экспериментальные и прикладные науки, практическая деятельность человека, рассматривающая повторяющиеся опыты как единое целое.

Мы познаём окружающий нас мир, выдвигая и проверяя по результатам эксперимента гипотезы о его свойствах. Получающиеся при этом выводы и заключения, никогда не обладая абсолютной достоверностью, тем не менее способны обострить интуицию исследователя, привести его от предварительных гипотез к более общим и строгим теориям. Другими словами, методы математической статистики — это мощный (а иногда, и единственный) многофункциональный инструмент в руках инженера и исследователя, медика и биолога, психолога и географа, студента и профессора.

Более ста лет тому назад человек, к своему изумлению, выяснил, что если бы случайность отсутствовала, ее нужно было бы изобрести, ибо, оказывается, в случайности заключены не только растерянность неопределенности, но и созидающая сила многовариантности. Как ни парадоксально, но поиск оптимальных условий протекания процесса наиболее эффективен, когда он реализуется в форме случайного поиска. Контроль качества изделий в производстве наиболее эффективен при обеспечении случайного отбора изделий из контролируемой партии. В конце концов, происхождению жизни, а, следовательно, и нашему существованию, мы обязаны случаю. Все мы помним пушкинскую строку „...гений, парадоксов друг“. Но, если внимательно дочитать поэта, то вслед за этой чудесной строкой следует не менее замечательная „...и случай, Бог — изобретатель“. Поэтому не следует считать математическую статистику только инструментом устранения досадного влияния случая. Нет, вместе с созидающим случаем математическая статистика является языком, на котором „Бог — изобретатель“ разговаривает с нами. И знать этот язык обязан каждый инженер, исследователь, каждый специалист. Если „перед ошибками захлопываем дверь, в смятеньи истина: „Как я войду теперь?“ — справедливо подметил Р. Тагор.

Практически любое решение, которое приходится принимать инженеру, руководителю производства, так или иначе требует применения методов обработки результатов наблюдений. Приведет ли внедренное новшество к повышению качества продукции? зависит ли наблюдаемый процесс от заданного фактора? существует

ли связь между исследуемыми величинами? сколь она сильна? — это типовые задачи математической статистики. Сколь долго можно ожидать безотказной работы прибора? как спланировать его заводские испытания? — это тоже задачи математической статистики. Перечисление таких примеров можно продолжать бесконечно, столь бесконечны и сложны взаимосвязи нашего разума с вечно меняющимся, бесконечным по форме и бездонным по содержанию миром, нас окружающим. Однако вместе с разумом человек обрел не только гордость бесконечности, но и привычку довольствоваться достаточным.

Почему появилась эта книга? Чем она отличается от других?

Можно надеяться, что, вняв изложенным аргументам, инженер или исследователь пожелает незамедлительно обострить свою интуицию, раздобыть чудодейственный инструмент решения своих повседневных задач.

Казалось бы, возможности прикладной математической статистики неоспоримы. Почему же тогда инженеры и ученые, мастера и рабочие остаются в неведении относительно этих возможностей? Имея многолетний опыт подготовки студентов старших курсов вузов, могу утверждать, что и сегодняшние выпускники имеют весьма смутное представление о современных методах математической статистики, чаще всего не способны применять их на практике. Преподавание этого курса в вузах находится далеко не на должном уровне. Практические навыки выпускников ограничиваются туманными воспоминаниями о различных определениях вероятности, иссушающими душу нудными задачами из теории вероятностей, заставляющими вычислять вероятности появления событий методами комбинаторики. Относительно математической статистики молодой специалист знает только одно — дело это темное и „лучшая из парадигм — это правило трех сигм“.

Для аспирантов и соискателей ученых степеней математическая статистика является чаще всего красивой рамкой или упаковкой для диссертации. Присутствуя на защитах диссертаций и оппонируя их, автор часто разочарованно убеждался, что соискатель не владеет основами обработки результатов наблюдений, а соответствующие разделы диссертации не более чем подарок коллеги-профессионала.

Так почему же инженеров и ученых промышленности не увлекают прелести прикладной статистики? Оставив в стороне мотивации, способные склонить их к этому, отметим, что они неизбежно столкнутся, прежде всего, с серьезнейшей проблемой — неимоверным количеством монографий, книг, статей, таблиц и справочников.

По данным российского исследователя А. И. Орлова, шестая часть публикаций в математике относится к теории вероятностей, математической статистике и их применениям в различных прикладных областях. Ежегодно появляются более 5 000 статей и книг по этой тематике (читатель вправе заметить: ну вот, еще одна появилась, но об этом позже). К настоящему времени известно более миллиона (!) работ по статистическим методам, причем только по прикладной статистике сохраняют актуальность более 100 000 статей и книг. В мире издаются более 400 журналов и периодических изданий по проблемам математической статистики.

Можно назвать десятки зарубежных журналов математической статистики для инженеров и исследователей: „Annals of the Institute of Statistical Mathematics“, „The Annals of Mathematical Statistics“ (AMS), „Journal of the Royal Statistical Society“ (JRSS), „Biometrics“, „Biometrika“, „Communication in Statistics“, „Journal of the American Statistical Association“ (JASA), „Technometrics“, „Statistica Neerlandica“, „Sunkhya“, „The American Statistical“, в то время как у нас в стране можно отметить лишь раздел „Математические методы испытаний“ в журнале „Заводская лаборатория“, журнал „Надежность и контроль качества“ (в 1991 году в нем выделено место для выпусков „Статистические методы“) и отчасти журнал „Теория вероятностей

и ее применения“ (большинство материалов которого не только начинающему, но и мне не по зубам).

Простая „лоция“ по океану публикаций в области математической статистики не поможет инженеру или исследователю, лишенному профессиональной подготовки в теоретических вопросах математической статистики. Да и вряд ли ему будут доступно большинство заокеанских книг и журналов — „материков и островов“ этого океана, о существовании которого подавляющая часть инженеров и исследователей и не догадывается.

Вряд ли еще в каких-нибудь дисциплинах можно встретить книги, содержащие такое огромное количество литературных ссылок (например, в книге Л. Закса „Статистическое оценивание“ — около 1500 ссылок, в книге М. Холлендера и Л. Вулфа „Непараметрические методы статистики“ — около 800). Даже разнобой в терминологии может привести неискушенного читателя в замешательство. Вряд ли он сразу сообразит, что „итерационный“ и „сериальный“ критерии, „фазочастотный“ и „знаковый сериальный“ критерии — это различные названия одного и того же критерия. А от такого, например, оборота лихого статистика-профессионала, как „проверив предварительно нормальность распределения критерием Лина–Мудхолкара, дополненным комбинированным критерием Шпигельхальтера, можно приступить к сравнению средних нескольких выборок, например, с помощью критерия Стьюдента–Ньюмена–Кейльса“, любой нормальный инженер неизбежно должен впасть в тихую грусть и навсегда потерять желание заглядывать в этот раздел науки.

Как же быть? Как сделать достоянием практикующих инженеров и исследователей сокровищницу наработанных мировой наукой эффективных методов статистического анализа результатов наблюдений? Эти методы адаптируются к реальным потребностям практиков и включаются в пособия, ориентированные на специалистов-нематематиков, как правило, через 10–15 лет после их появления.

Такой попыткой является предлагаемая книга. Математическая статистика — наука, устроенная довольно своеобразно, и ее применение — искусство, требующее не только знаний, но и практики, опыта, чутья и интуиции. Сделать такую практику достоянием инженеров и исследователей — цель автора книги. Автор стремился прежде всего отобразить богатую палитру методологических подходов прикладной математической статистики. Он исходил из того, что инженеров и научных работников нужно знакомить с новыми методами и удачными приемами, если даже их приходится излагать на эвристическом уровне. Цель книги — научить пользоваться прикладными методами математической статистики, не требуя владения ее теоретическими основами. Недостаток понимания основ восполняется подробными рекомендациями и предостережениями, а также обширнейшей библиографией. Древние греки справедливо заметили, что „способности чахнут и теряют естественность при соприкосновении с иссушающими природу учеными наставлениями“. Поэтому автор стремился елико возможно избегать таковых и, следуя мудрости Ньютона — „при изучении наук примеры полезнее правил“, обратился к наиболее эффективному для пользователя методу изложения материала — в форме демонстрационных примеров и задач. Изложение техники и последовательности расчетных процедур заменены, там, где это не в ущерб истине, пояснением конкретных примеров. При этом автор исходил из того, что заинтересованный только приложениями математической статистики инженер-практик не будет читать все подряд, а попытается разыскать пример, похожий на тот, что его интересует. И он будет прав, ибо прикладная математическая статистика не является наукой, которая может быть изучена только путем чтения. Умение и своеобразное чутье выбрать правильный метод приходит только в процессе решения практических задач. Следуя мудрецу Дейлу Карнеги, признаем, что „в сущности, всё, в конечном счете, сводится к одному — нужно

практиковаться, практиковаться и практиковаться". Компьютер поможет выполнить расчеты, но не заменит ни процесс формирования гипотезы, ни творчество, проистекающее из воображения.

В книге предлагаются около 400 статистических приемов решения всевозможных практических задач. То, что автор приводит такое изобилие различных современных методов, многие из которых известны пока не каждому профессионалу, не является следствием его неразборчивости, а делается им осмысленно.

Причин тому несколько:

— современные компьютеры сделали доступными ранее недоступные в вычислительном отношении методы;

— не следует исключать стремление читателя к разнообразию, проистекающее от природного любопытства;

— демонстрация различных подходов к одной и той же задаче помогает глубже ее осмыслить.

Автор на себе опробовал этот путь, именно так начав знакомство с прикладной математической статистикой, будучи молодым специалистом-электронщиком. Конечный итог даже только знакомства со всем разнообразием методов — путь к пониманию математической статистики, к овладению практическими навыками пользования ее методами. Ведь каждый метод — это демонстрация тех или иных положений статистики.

Обилие различных критериев и оценок в статистике не должно пугать — это следствие множественности ситуаций, возникающих в реальной жизни. Знакомство с массой различных приемов статистической обработки данных способствует демонстрации самого механизма, способа мышления, методологии прикладной статистики, помогает глубокому усвоению как ее методов, так и философии. Пользователя не должна смущать эквивалентность некоторых методов, как не пугает его возможность решать одну и ту же жизненную коллизию разными способами. Знакомство с множеством практических приемов позволяет пользователю почувствовать „воздух" математической статистики, на уровне подсознания уяснить ее методологию, внутреннюю логику, разнообразие и остроумие подходов. Это способ учиться статистике, да и необходимость выбора подходящего приема из большой совокупности возможных — неплохой тренинг на долгом пути знакомства с математической статистикой для любого инженера или исследователя. Так или иначе, пользователю предлагается совокупность методов и приемов прикладной математической статистики, которая никогда еще не собиралась в одной книге.

Изложение материала в книге преследует цель дать по каждому методу набор стандартной информации, включающий в себя: логическое обоснование, связь с другими методами, назначение, авторство, методику расчетов, необходимые таблицы и аппроксимации, указания по особенностям применения и статистическим характеристикам, набор демонстрационных задач. Располагая такой исчерпывающей информацией, пользователь может решать свои практические задачи, не обращаясь ни к каким другим дополнительным источникам.

Исчерпывающий характер предлагаемой информации дает пользователю уникальный шанс приобрести в компактной форме весь мировой опыт прикладной математической статистики, в десятки раз сократить время поиска необходимых методов в океане публикаций, в сотни раз повысить эффективность своей работы за счет принятия оптимальных решений в море возможных, сокращения объема испытаний и экспериментальных работ, повышения качества продукции. Эта книга полезна для аспирантов и ученых, желающих (и обязанных) на современном уровне проводить эксперименты и обрабатывать их результаты. Она может быть использована студентами вузов как универсальное справочное пособие.

Что есть в книге и немного теоретических основ статистики

Предлагаемая книга включает в себя пять глав. Все главы связаны между собой единой логикой методологического подхода и посвящены отдельным важнейшим, имеющим самостоятельное практическое значение задачам прикладной математической статистики.

Состав и тематика отдельных глав последовательно отражают логику развития самой теории вероятностей и математической статистики. Мы уже говорили о том, что потребность каким-то образом справиться с потоком информации у человека возникла давно, сразу после того, как он стал понимать, что это информация. Естественно, первое, что отметит каждый из нас — это наличие в потоке информации определенных закономерностей. Источником такой закономерности является нечто объективное, содержащееся в природе наблюдаемого процесса или явления. Проявлением этого „нечто" является частота появления определенных событий (величин). Мы отмечаем, что одни события встречаются чаще (или реже), чем другие. Это наблюдение потребовало найти „нечто", которое, говоря словами Гете, „единичное искусно обобщает, объединяя все в торжественный аккорд", и это „нечто" было названо *вероятностью.* Обращу внимание читателя на одну любопытную деталь. Хорошо известны такие журналы по прикладной математической статистике, как „Технометрикс", „Биометрика", „Эконометрика", „Психометрика". Казалось бы, они являются журналами для специалистов разных профессий, но все они — журналы прикладной математической статистики, и именно она — математическая статистика — является той „метрикой", которая объединяет все прикладные науки, имеющие дело с потоком числовых данных, ибо, как тонко заметил Валлювар, „что, в сущности, буква и цифра? Не глаза ли два, которым открыта вся суть естества?"

Вернемся к вероятности. Понимая под ней частоту появления отдельных событий в наблюдаемом потоке, человек, естественно, попытался связать вероятность появления случайного события с ее количественным значением. Такая связь может быть описана как *функция распределения вероятностей* или *плотность распределения вероятностей*, являющаяся производной от функции распределения. Описанию распределений вероятностей посвящена первая глава книги.

У читателя может возникнуть вопрос: а зачем, собственно, специально описывать какие-то функции, и откуда они вообще взялись? В конце концов, существуют достаточно хорошо разработанные в математическом анализе приемы исследования функций, изучаемые еще в школе. Можно, конечно, воспользоваться этими приемами, но мы все-таки имеем дело не просто с математическим соотношением, а с вероятностной моделью. Поэтому для описания функций распределения вероятностей используются специальные меры (параметры), называемые *моментами.* Отдельные моменты и их комбинации характеризуют такие важные в практическом отношении характеристики распределения вероятностей, как центр группирования данных, степень их рассеяния относительно центра, поведение случайных величин в районе центра группирования, симметричность распределения вероятностей.

Среди известных законов распределения вероятностей каждый наверняка назовет *нормальный закон.* Но немногие спросят: а почему он называется нормальным? Уж не потому ли, что другие распределения ненормальны? Вопрос не так наивен, как кажется. Ничего нормального, судя по математической функции плотности распределения вероятностей

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\},$$

в нем нет, если не считать, что она содержит достаточно полный набор символов, обычно употребляемых в математике.

Дело в том, что если наблюдаемый процесс зависит от совокупности большого числа взаимонезависимых (или слабозависимых) факторов, вклад каждого из которых в процесс мал, то, поверь мне, читатель, на слово (это доказали весьма ученые мужи), мы неизбежно придем к нормальному распределению вероятностей. Наверное, оттого, что описанная ситуация представляется нам нормальной, назван нормальным и закон, отражающий влияние случайных факторов на результирующий процесс.

Большинство используемых в прикладной математической статистике распределений вероятностей, так или иначе связано с нормальным законом или сводится к нему. В первой главе книги „Распределения вероятностей случайных величин" приведены подробные сведения более чем о 30 непрерывных и дискретных распределениях вероятностей, исчерпывающих большинство мыслимых практических ситуаций. Описаны области их применения, методы расчета вероятностей и необходимые аппроксимации.

Во второй главе книги „Оценка параметров распределений вероятностей" рассмотрены методы оценки параметров различных распределений вероятностей. Математическое понятие *параметр распределения* в практике инженера представлен такими знакомыми категориями, как средняя наработка на отказ, интенсивность отказов, точность показаний измерительного прибора и т. п. Поэтому методы оценки параметров распределений для инженера и исследователя являются методами извлечения из результатов наблюдений, испытаний и экспериментов информации, позволяющей оценить качество изделий, уровень принятых технических решений.

Рассмотрены точечные и интервальные оценки всех наиболее распространенных распределений. Большое внимание уделено „быстрым" упрощенным оценкам, а также *непараметрическим* (свободным от распределения) и так называемым *робастным* (robust) *оценкам*, устойчивым к засорению выборок посторонними наблюдениями.

Для практического применения методов математической статистики чрезвычайно важно знание закона распределения вероятностей. По существу, сама изучаемая величина представлена для исследователя только законом распределения вероятностей реализации ее значений.

Попытка применить методы анализа результатов наблюдений, разработанные для конкретных законов распределения вероятностей, в условиях, когда реальное распределение отличается от гипотетического, является самой распространенной ошибкой, приводящей к неверным выводам и, в конечном итоге, к существенным материальным потерям и затратам времени. Именно поэтому любая обработка результатов наблюдений должна неизменно начинаться с ответа на главный вопрос: каким законом описывается распределение вероятностей совокупности, из которой извлечена обрабатываемая выборка случайных величин? На практике обычно эта проблема формулируется следующим образом. Выдвигается гипотеза, утверждающая, что наблюдаемое распределение случайных величин описывается конкретным законом (нормальным, экспоненциальным, равномерным, Вейбулла и т. д.). Задача первичного анализа — принять или отклонить выдвинутую гипотезу. Если ни одна из гипотез относительно формы закона распределения не принимается, то может быть сформулирована более мягкая гипотеза, например — „наблюдаемое распределение вероятностей симметрично относительно определенной точки". Установление даже этого факта дает в руки исследователя более эффективные методы анализа наблюдений, чем при полном незнании закона распределения вероятностей. И, наконец, если исследователь не получил достаточных оснований для выбора типа распределения, то возникает задача подбора кривой распределения непосредственно по экспериментальным данным.

Критерии проверки гипотез о законе распределения вероятностей принято называть *критериями согласия*, подразделяя их на две группы — общие и специальные критерии согласия. *Общие критерии согласия* применимы к формулировке гипотезы о согласии наблюдаемых результатов с любым априорно предполагаемым распределением. *Специальными критериями* проверяются специальные гипотезы, формулирующие согласие с конкретной формой распределения — нормальной, экспоненциальной и т. п. Такие критерии носят соответствующие названия — *критерии нормальности*, *критерии экспоненциальности*, *критерии равномерности.* Естественно, что при формулировании специфичных требований общие критерии согласия могут быть трансформированы в специальные критерии.

Следует отметить, что многообразие возможных альтернатив, противостоящих выдвинутой гипотезе, порождает и чрезвычайное многообразие статистических критериев, имеющих различную мощность по отношению к различным альтернативам. В третьей главе „Методы анализа законов распределения вероятностей случайных величин“ представлена широкая гамма критериев согласия (более 80), впервые собранная в одной книге.

В четвертой главе „Проверка гипотез о значениях параметров распределений вероятностей“ рассматриваются методы проверки предположений о значениях параметров распределений. Формулирование гипотез о свойствах окружающего нас мира и проверка их непосредственно наблюдениями или с помощью целенаправленного экспериментирования составляет основу того, чтó мы называем наукой или научной деятельностью.

Проверка гипотез применительно к потребностям ежедневной практики инженера или исследователя приобретает вполне конкретный смысл, зависящий от специфики наблюдаемых реалий, особенностей возникающих коллизий, потребностей практики, стимулировавших сам процесс зарождения и формулирования гипотезы. Например, часто встречающаяся задача — соответствуют ли параметры разработанного изделия предъявляемым требованиям — в математико-статистической формулировке может звучать так: „необходимо проверить гипотезу о том, что параметр ξ распределения случайной величины X превосходит заданную контрольную величину ξ_0“.

Сколь разнообразен и сложен окружающий нас мир, столь велико и разнообразно семейство возможных гипотез о его свойствах. Поэтому четвертая глава является самым объемным разделом книги.

Автор надеется, что знакомство с многочисленными примерами решения задач является лучшим способом узнавания палитры методов прикладной статистики. Впрочем, в случае нежелания наблюдать палитру прагматик, располагая этой книгой, получает инструмент решения своих практических задач без необходимости вникать в глубину захватывающего мира обработки результатов наблюдений. Следует всегда помнить об антагонистическом противоречии между категоричностью и надежностью высказывания по гипотезе: надежное высказывание некатегорично, категоричное высказывание ненадежно. Мы выдвигаем гипотезу и отвергаем ее тогда, когда получаем результат, маловероятный при истинности выдвинутой гипотезы. Принятая граница маловероятности называется *уровнем значимости.* Если мы наблюдаем событие, вероятность появления которого не превышает уровень значимости, мы называем его *значимым*, используя этот термин в качестве антонима термину *случайность.* Такой метод принятия решений в статистике получил название *принципа значимости.*

При проверке гипотез возможны ошибки двух типов — неправильное отклонение верной гипотезы (ошибка первого рода) и неправильное принятие ложной гипотезы (ошибка второго рода). Следует помнить, что уровень значимости должен устанавливаться перед получением данных. Это требование для практи-

ка вообще является некоторой головоломкой, но, уважаемый читатель, поверь, что это очень важно. Ты можешь задать естественный вопрос: а какой уровень значимости или достоверности следует выбирать? Увы, но это твоя проблема, читатель, а не математической статистики. Достоверность, с которой ты бы хотел получить ответ на поставленный вопрос, должна определяться тобой исходя из практической ситуации. Желание повысить достоверность заключения по гипотезе связано с увеличением затрат, стремление снизить затраты на проверку гипотезы неизбежно приводит к повышенному риску принять ложное решение. Всё должны определять конкретная ситуация и цена риска. Например, проектируя атомный реактор или переходя дорогу перед транспортом, мы стремимся свести к нулю риск даже повышением затрат, ибо цена высока. Одним из методов сохранения достоверности выводов при снижении затрат на проверку гипотезы является выбор эффективного статистического приема обработки результатов наблюдений. В этом поможет предлагаемая книга.

Обычно на практике применяются уровни значимости 0,01; 0,05; 0,1. Важно неукоснительно выполнять основное требование — гипотезы должны быть выдвинуты перед статистическим анализом, а сам числовой материал не должен быть использован для выдвижения гипотезы. Гипотезы, выдвинутые на основе анализа полученного материала, могут быть полезны только в качестве новых гипотез для последующих проверок. Перефразируя Томаса Гексли, укажем на великую трагедию математической статистики как науки — „она способна уничтожить прекрасную гипотезу одним безобразным фактом“.

В первых четырех главах книги рассмотрены методы и приемы математической статистики, позволяющие оценить параметры статистических совокупностей, сравнить их между собой. При этом, как правило, предполагалась взаимная независимость сравниваемых совокупностей. В последней, пятой главе „Методы исследования связей между случайными величинами“ рассмотрены вопросы оценки связей между статистическими совокупностями. В ней последовательно излагаются методы и приемы дисперсионного, корреляционного и регрессионного анализов, являющихся последовательными ступенями при изучении характера и особенностей связей между случайными величинами.

Методами дисперсионного анализа устанавливается влияние заданного фактора на процесс, отображаемый наблюдаемой статистической совокупностью данных. Корреляционный анализ позволяет оценить силу (степень) такой связи, а методами регрессионного анализа можно установить конкретную математическую модель, адекватно отражающую установленную связь.

Стоит ли иметь эту книгу? Подумай!

Здесь будет немного рекламы. Но кто без нее сегодня обойдется? Если ты приобретешь эту книгу, ты станешь обладателем системы знаний в области обработки и анализа результатов наблюдений.

У читателя может возникнуть вопрос: а не проще ли воспользоваться одним из пакетов статистического программного обеспечения и поручить ему обработку результатов наблюдений, оставив за собой роль беспечного наблюдателя? Несомненно, такой пакет является изящным и мощным инструментом в руках профессионала, которому знаком язык математической статистики, но он вряд ли поможет инженеру или исследователю, далекому от мира формальной математики и от терминологического языка математической статистики, творчески осмыслить систему решения тех или иных задач, логику изучения процессов методами статистики. Предлагаемая книга может с успехом обеспечить начальный тренинг неподготовленному пользователю, подготовив тем самым ему плацдарм для штурма высот современного программного статистического обеспечения.

В книге содержится подробное описание около 400 задач, иллюстрирующих решение практически любой проблемы обработки результатов наблюдений. Перевод методов математической статистики, иногда довольно сложно описываемых языком формальной математики, на язык последовательности выполнения элементарных вычислительных операций делает их доступными неподготовленному пользователю, позволяет ему успешно справляться с возникающими затруднениями на высоком уровне профессионала, не требуя его квалификации и подготовки.

Более 200 математико-статистических таблиц являются мощной „базой данных“, они предоставляют в распоряжение читателя практически весь арсенал известных таблиц, бо́льшая часть которых пока содержится только в специализированных научных журналах, мало известных большинству инженеров-практиков. В распоряжение читателя, который пожелает углубиться в основания прикладной математической статистики, предлагается обширный перечень литературы, содержащий около 700 наименований.

Прикладная статистика становится таковой тогда, когда ее методы широко применяются на практике. Наша цель — сделать ее таковой, доведя методы математической статистики до тех, кто, пасуя перед ее высокой математизацией, не вникал в достаточно простой смысл математических символов.

Некоторое представление об объеме сведений, содержащихся в книге, и ее базе данных дают содержание глав и перечень основных использованных информационных источников. Весьма полезен также прилагаемый перечень задач, являющийся ориентиром при выборе задачи, похожей на ту, которую предстоит решить пользователю.

С уважением,

Доктор технических наук,
профессор
А. И. Кобзарь

ГЛАВА 1

РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ СЛУЧАЙНЫХ ВЕЛИЧИН

Общие положения. Распределение значений случайной величины по вероятности их появления характеризуется интегральной функцией распределения (законом распределения) и плотностью вероятностей.

Интегральная функция (закон) распределения случайной величины X обозначается через $F(x)$ и определяется как

$$F(x) = \mathbf{P}[X \leqslant x],$$

где $\mathbf{P}[X \leqslant x]$ — вероятность того, что значение случайной величины X не превысит x.

Плотность вероятности выражается формулой

$$f(x) = \frac{dF(x)}{dx}.$$

Свойства функций $F(x)$ и $f(x)$ определяются соотношениями:

$$F(x) = \int\limits_{-\infty}^{x} f(x)\,dx, \quad f(x) \geqslant 0,$$

$$F(-\infty) = \lim_{x\to-\infty} F(x) = 0, \quad \int\limits_{-\infty}^{\infty} f(x)\,dx = 1,$$

$$F(+\infty) = \lim_{x\to+\infty} F(x) = 1, \quad \mathbf{P}[x_1 < x \leqslant x_2] = F(x_2) - F(x_1),$$

$$F(x_2) \leqslant F(x_1), \quad \text{если} \quad x_2 \leqslant x_1.$$

Если известен закон распределения вероятностей случайной величины, то можно решать многие практические задачи, возникающие при статистическом анализе экспериментальных данных. Например, вычисление вероятности попадания случайной величины в заданный интервал, вероятности превышения случайной величиной заданного значения и т. п. Определенное представление о функции распределения дают его *квантили*. По определению $\alpha \cdot 100\%$-я квантиль (α-квантиль) распределения, обозначаемая x_α, соответствует условию

$$\mathbf{P}(X \leqslant x_\alpha) = F(x_\alpha) = \alpha.$$

Для описания функций распределения пользуются специальными мерами, позволяющими охарактеризовать положение, форму и другие их особенности.

Центр распределения характеризуется средним значением μ, медианой **Me**, модой **Mo**. *Среднее значение* μ равно первому начальному моменту, *медиана* является 50%-й квантилью распределения, *мода* соответствует значению $\widetilde{x}$ случайной величины, для которого $f(\widetilde{x}) = \max$.

Рассеяние случайных величин вокруг центра группирования оценивается *дисперсией* (математическое ожидание квадрата отклонения случайной величины от

среднего значения), *стандартным отклонением* (квадратный корень из дисперсии), *коэффициентом вариации* (отношение стандартного отклонения к математическому ожиданию случайной величины), *размахом* $w = x_{\max} - x_{\min}$.

Симметричность распределения характеризуется *коэффициентом асимметрии* α_3, особенности поведения случайной величины в области максимума ее плотности описываются *коэффициентом эксцесса* α_4.

Подробно с теоретическими основами теории вероятностей и функциями распределения можно ознакомиться, обратившись к монографиям [1–22]. Исходя из природы возникновения случайных величин, различают распределения непрерывных и дискретных случайных величин.

1.1. Непрерывные распределения

1.1.1. Нормальное распределение

Описание, применение. Наиболее широко применяемое распределение. В самом названии распределения отражена идея его универсальности. Нормальное распределение подробно исследовано в работах Муавра, Лапласа, Гаусса, Чебышева, Ляпунова, Бернштейна. Теоретической основой нормального закона распределения вероятностей является центральная предельная теорема Ляпунова, утверждающая, что распределение суммы независимых случайных величин с любым исходным распределением будет нормальным, если число слагаемых достаточно велико, а вклад каждого в сумму мал.

Нормальное распределение является краеугольным камнем математической статистики в силу ряда причин:

— схема его возникновения соответствует многим реальным физическим процессам, порождающим результаты обрабатываемых наблюдений;

— при возрастании объема выборки предельное распределение для большинства распределений является нормальным и с успехом может использоваться для аппроксимации последних;

— нормальное распределение обладает рядом благоприятных математико-статистических свойств (легко нормируется и аппроксимируется, обладает свойством аддитивности).

В теории надежности нормальное распределение обычно используется для описания износовых отказов, интенсивность которых со временем возрастает.

Свойства

Обозначение	$N(\mu, \sigma)$
Параметры	μ, σ
Плотность	$f(x; \mu, \sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right\}$
Функция распределения	$F(x; \mu, \sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left\{-\dfrac{1}{2}\left(\dfrac{t-\mu}{\sigma}\right)^2\right\} dt$ $(-\infty < x < +\infty)$
Среднее	$\mathbf{M}(x) = \mu$
Дисперсия	$\mathbf{D}(x) = \sigma^2$
Стандартное отклонение	σ
Коэффициент вариации	$v = \dfrac{\sigma}{\mu}$

Коэффициент асимметрии	$\alpha_3 = 0$
Коэффициент эксцесса	$\alpha_4 = 3$
Мода	$\mathbf{Mo} = \mu$
Медиана	$\mathbf{Me} = \mu$

Распределение симметрично относительно точки $x = \mu$ и имеет два параметра μ и σ, совпадающих со средним значением и стандартным отклонением.

Для удобства в практических приложениях применяется нормированная случайная величина $z = (x - \mu)/\sigma$, распределение которой называется *стандартным нормальным* с нулевым средним и единичной дисперсией: $N(0,1)$.

В большинстве пособий по теории вероятностей и математической статистике приводятся таблицы функции

$$F(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} \exp\left(-\frac{z^2}{2}\right) dz$$

или связанной с ней функции (интеграла) Лапласа

$$\Phi(z) = \sqrt{\frac{2}{\pi}} \int\limits_{0}^{z} \exp\left(-\frac{z^2}{2}\right) dz = 2F(z) - 1.$$

Очевидно, что

$$F(-z) = 1 - F(z) \quad \text{и} \quad \mathbf{P}(a \leqslant z \leqslant b) = F(b) - F(a).$$

Сумма $y = \sum\limits_{i=1}^{n} a_i x_i$ нормально распределенных случайных величин $N(\mu_i; \sigma_i)$ распределена нормально. p-квантиль нормально распределенной случайной величины $N(\mu, \sigma)$ связан с квантилью u_p^c случайной величины, имеющей стандартное нормальное распределение $N(0,1)$, соотношением

$$u_p = \mu + u_p^c \sigma.$$

В силу симметричности нормального распределения $u_p^c = -u_{1-p}^c$.

Обширные таблицы нормального распределения приведены в [23–29]. Часть их воспроизведена в табл. 1. Так как таблицы всегда ограничены, приведем известные аппроксимации, достаточно легко реализуемые применением существующих массовых микрокалькуляторов или персональных ЭВМ.

Аппроксимация 1 [30]:

$$F(z) = 1 - \left(\sqrt{2\pi} e^{z^2/2}\right)^{-1} \sum_{i=1}^{3} a_i \lambda^i,$$

где $\lambda = (1 + bz)^{-1}$; $b = 0{,}33267$; $a_1 = 0{,}4361836$; $a_2 = -0{,}1201676$; $a_3 = 0{,}937298$.

Абсолютная погрешность $\leqslant 1 \cdot 10^{-5}$.

Аппроксимация 2 [30]:

$$F(z) = 1 - \left(\sqrt{2\pi} e^{z^2/2}\right)^{-1} \sum_{i=1}^{5} a_i \lambda^i,$$

где $\lambda = (1 + bz)^{-1}$; $b = 0{,}2316419$; $a_1 = 0{,}31938153$; $a_2 = -0{,}35656378$; $a_3 = 1{,}7814779$; $a_4 = -1{,}821256$; $a_5 = 1{,}3302744$.

Абсолютная погрешность $\leqslant 1 \cdot 10^{-7}$.

Таблица 1

Квантили стандартного нормального распределения

p	u_p	p	u_p	p	u_p
0,50	0,00000	0,820	0,91536	0,995000	2,57583
0,52	0,05015	0,840	0,99446	0,996000	2,65207
0,54	0,10043	0,860	1,08032	0,997000	2,74778
0,56	0,15097	0,880	1,17499	0,998000	2,87816
0,58	0,20189	0,900	1,28155	0,999000	3,09023
0,60	0,25335	0,910	1,34075	0,999200	3,15591
0,62	0,30548	0,920	1,40507	0,999400	3,23888
0,64	0,35846	0,930	1,47579	0,999500	3,29053
0,66	0,41246	0,940	1,55477	0,999600	3,35279
0,68	0,46770	0,950	1,64485	0,999700	3,43161
0,70	0,52440	0,960	1,75069	0,999800	3,54008
0,72	0,58284	0,970	1,88079	0,999900	3,71902
0,74	0,64334	0,980	2,05375	0,999950	3,89059
0,76	0,70630	0,990	2,32635	0,999990	4,26489
0,78	0,77219	0,992	2,40891	0,999995	4,41717
0,80	0,84162	0,994	2,51214	0,999999	4,75342

Аппроксимация 3 [31–33]:

$$\Phi(z) = 1 - \left(1 + 10^{-6} z(c_6 + z(c_5 + z(c_4 + z(c_3 + z(c_2 + zc_1)))))\right)^{-16},$$

где $c_1 = 5{,}383$; $c_2 = 48{,}891$; $c_3 = 38{,}004$; $c_4 = 3277{,}626$; $c_5 = 21141{,}006$; $c_6 = 49867{,}347$.

Абсолютная погрешность $\leqslant 5 \cdot 10^{-7}$.

Аппроксимация 4 [32, 34]:

$$\Phi(z) = \sqrt{\frac{2}{\pi}} \sum_{i=0}^{4} \frac{(-1)^i z^{2i+1}}{2^i i!(2i+1)}.$$

Относительная погрешность $\leqslant 0{,}3\%$.

Аппроксимация 5 [32]:

$$\Phi(z) = \left\{1 - \exp\left(-\frac{2z^2}{\pi}\right)\right\}^{\frac{1}{2}}, \quad 0 < z \leqslant 1{,}96; \quad 0 \leqslant \Phi(z) \leqslant 0{,}95.$$

Абсолютная погрешность $\leqslant 0{,}114$.

Аппроксимация 6 [32]:

$$\Phi(z) = \left\{1 - \left(1 + kz^2\right)^{-1} \exp\left[\left(\frac{2}{\pi} - k\right)\left(-z^2\right)\right]\right\}^{\frac{1}{2}},$$

$$0 < z \leqslant 5{,}327; \quad 0 \leqslant \Phi(z) \leqslant 1 - 1 \cdot 10^{-7},$$

где $k = 0{,}1253$.

Относительная погрешность $\leqslant 1\%$.

Аппроксимация 7 [35]:

$$F(z) = 1 - \frac{1}{2}\left(1 + a_1\frac{z}{\sqrt{2}} + a_2\frac{z^2}{2} + a_3\frac{z^3}{2\sqrt{2}} + a_4\frac{z^4}{4}\right)^{-4},$$

где $a_1 = 0{,}278393$; $a_2 = 0{,}230389$; $a_3 = 0{,}0000972$; $a_4 = 0{,}078108$.

Аппроксимация 8 [36]:

$$F(z) = 1 - 0{,}852 \exp\left\{-\left(\frac{z + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\}, \quad z > 0;$$

при $z < 0 \quad F(-z) = 1 - F(z)$.

Абсолютная погрешность $\leqslant 0{,}0002$.

Аппроксимация 9 [37]:

$$u_p = -\left[-2\ln(\sqrt{2\pi}p)\right]^{\frac{1}{2}}, \quad p \leqslant 0{,}01.$$

Аппроксимация 10 [38]:

$$u_p = t - \frac{\sum\limits_{i=0}^{2} c_i t^i}{1 + \sum\limits_{i=1}^{3} d_i t^i}; \quad t = [-2\ln(1-p)]^{\frac{1}{2}}.$$

где $c_0 = 2{,}515517$; $c_1 = 0{,}802853$; $c_2 = 0{,}010328$; $d_1 = 1{,}432788$; $d_2 = 0{,}189269$; $d_3 = 0{,}001308$.

Абсолютная погрешность $\leqslant 0{,}00045$.

Аппроксимация 11 [37]:

$$u_p = -\left\{-2\ln\left[\sqrt{2\pi}p(-2\ln\left(\sqrt{2\pi}p)^{\frac{1}{2}}\right)\right]\right\}^{\frac{1}{2}}, \quad p \leqslant 0{,}05.$$

Аппроксимация 12 [32]:

$$u_{\frac{1+p}{2}} = \left[-\frac{\pi}{2}\ln\left(1 - p^2\right)\right]^{\frac{1}{2}}, \quad p < 0{,}99.$$

Относительная погрешность $\leqslant 3{,}7\%$.

Аппроксимация 13 [32]:

$$u_{\frac{1+p}{2}} = \sqrt{\frac{\pi}{2}}\left[-\ln(1 - p^k)\right]^{\frac{1}{k}}, \quad 0 < p \leqslant 1 - 1 \cdot 10^{-7},$$

где $k = 1{,}898$.

Относительная погрешность $\leqslant 2\%$.

Аппроксимация 14 [32]:

$$u_{\frac{1+p}{2}} = \left\{\pi\left\{\frac{1}{41}\left[-\ln(1 - p^4)^{\frac{3}{2}}\right] - \ln\sqrt{1 - p^2}\right\}\right\}^{\frac{1}{2}}, \quad 0 < p \leqslant 1 - 1 \cdot 10^{-7}.$$

Относительная погрешность $\leqslant 0{,}3\%$.

Аппроксимация 15 [39]:

$$u_p = 4{,}91\left\{p^{0{,}14} - (1-p)^{0{,}14}\right\}.$$

Относительная погрешность $\leqslant 0{,}3\%$.

Аппроксимация 16 [36]:

$$u_p = 2{,}0637\left(\ln\frac{1}{1-p} - 0{,}16\right)^{0{,}4274} - 1{,}5774, \quad 0{,}5 \leqslant p \leqslant 0{,}999.$$

Абсолютная погрешность $\leqslant 0{,}0008$.

Аппроксимация 17 [33]:

$$u_p = \left[\pi\left\{\frac{1}{23}\left[-\ln(1-p)^{\frac{5}{4}}\right] - \ln\sqrt{2(1-p)}\right\}\right]^{\frac{1}{2}},$$

$$0{,}0455 > 1-p > 1{,}244\cdot 10^{-5};\quad 2 \leqslant u_p \leqslant 8.$$

Относительная погрешность $\leqslant 0{,}6\%$.

Аппроксимация 18 [40]:

$$F(z) = 1 - \frac{1}{2}\exp(bz + az^2),\quad z > 0,$$

где $b = -0{,}717$; $a = -0{,}416$.

Аппроксимация 19 [40]:

$$u_{\frac{1+p}{2}} = \frac{1}{2a}\left(-b - \sqrt{b^2 + 4a\ln(1-p)}\right),\quad p > 0{,}5,$$

где $b = -0{,}717$; $a = -0{,}416$.

Относительная погрешность $\leqslant 0{,}5\%$.

Аппроксимация 20 [114]:

$$F(z) = \left\{1 + \exp\left[-0{,}0725z\left(22 + z^{1{,}96}\right)\right]\right\}^{-1},\quad z \geqslant 0;$$

при $z < 0$ $\quad F(z) = 1 - F(-z)$.

Точность при трех знаках абсолютная, в четвертом знаке максимальная погрешность 0,02%. Это одна из наиболее простых и весьма точных аппроксимаций.

Аппроксимация 21 [114]:

$$u_p = \frac{1{,}24 + 0{,}85K^{0{,}657}}{1 + 0{,}0001K^3 + \dfrac{2{,}38}{K}},\quad p \geqslant 0{,}5,$$

где $K = -\ln\left(\dfrac{1}{p} - 1\right)$.

Относительная погрешность 0,3%.

Для многих приложений требуется знание математического ожидания i-й порядковой статистики в выборке объема n из стандартного нормального распределения, т. е. математического ожидания i-го по номеру члена выборки, упорядоченной по возрастанию.

Известно, что математическое ожидание i-го по величине члена выборки из стандартного нормального распределения $\mathbf{M}(z_i)$ равно квантили $u_{p'}$, где $p' = \dfrac{i - 3/8}{n + 1/4}$ [40].

Таким образом, аппроксимации для квантилей могут быть использованы и для аппроксимации математических ожиданий порядковых статистик, что позволяет отказаться от применения громоздких таблиц.

Поясним применение предложенных аппроксимаций при решении практических задач.

Задача 1. *Вычислить значение функции распределения вероятностей в точке $x = 200$, если случайная величина распределена нормально со средним $\mu = 100$ и дисперсией $\sigma^2 = 2500$.*

Значение нормированной переменной определим по формуле

$$z = \frac{x-\mu}{\sigma} = \frac{200-100}{50} = 2.$$

Таким образом, смысл задачи состоит в нахождении значения $F(z) = \mathbf{P}(z \leqslant 2)$.

Для вычисления $F(2)$ воспользуемся приведенными ранее аппроксимациями. Для контроля будем иметь в виду, что точное табличное значение $F(2) = 0{,}97725$.

Аппроксимация 1. Вычисляем $\lambda = \dfrac{1}{1 + 0{,}332567 \cdot 2} = 0{,}60047798$.

Далее имеем $\left(\sqrt{2\pi}e^{\frac{z^2}{2}}\right)^{-1} = \left(2{,}506628275e^2\right)^{-1} = 0{,}053990966$.

Находим $\sum\limits_{i=1}^{3} a_i\lambda^i = 0{,}436836 \cdot 0{,}60047798 - 0{,}1201676 \cdot (0{,}60047798)^2 + \ldots = 0{,}421529962$.

Вычисляем $F(2) = 1 - 0{,}053990966 \cdot 0{,}421529962 = 0{,}97724119$.

Относительная погрешность $\delta = \dfrac{0{,}97725 - 0{,}97724119}{0{,}97725} = 0{,}0009\%$.

Аппроксимация 2. Вычисляем

$$\lambda = \frac{1}{1 + 0{,}2316419 \cdot 2} = 0{,}683394431 \quad \text{и} \quad \left(\sqrt{2\pi}\exp\left(\frac{z^2}{2}\right)\right)^{-1} = 0{,}053990966.$$

Вычисляем $\sum\limits_{i=1}^{5} a_i \cdot \lambda^i = 0{,}31938153 \cdot 0{,}683394431 - 0{,}35656378 \cdot (0{,}683394431)^2 + \ldots = 0{,}421367913$.

Находим $F(2) = 1 - 0{,}053990966 \cdot 0{,}421367913 = 0{,}977249939$.

Относительная погрешность $\delta = \dfrac{0{,}97725 - 0{,}9772499}{0{,}97725} = 0{,}000006\%$.

Аппроксимация 3. Последовательно вычисляем $5{,}383 \cdot 2 + 48{,}891 = 59{,}657$; $59{,}657 \cdot 2 + 38{,}004 = 157{,}318$; $157{,}318 \cdot 2 + 3277{,}626 = 3592{,}262$; $3592{,}262 \cdot 2 + 21141{,}006 = 28325{,}53$; $28325{,}5 \cdot 2 + 49867{,}347 = 1065518{,}407$; $(1 + 10^{-6} \cdot 2 \cdot 1065518{,}407)^{-16} = 0{,}45500513$.

Находим $\Phi(2) = 1 - 0{,}45500513 = 0{,}954499486$ и окончательно

$$F(2) = \frac{1 + \Phi(2)}{2} = \frac{1 + 0{,}954499486}{2} = 0{,}977249743.$$

Относительная погрешность $\delta = 0{,}00003\%$.

Аппроксимация 4. Вычисляем

$$\sum_{i=0}^{4} \frac{(-1)^i \cdot z^{2i}}{2^i \cdot i! \cdot (2i+1)} = 1 - \frac{2^2}{2 \cdot 3} + \frac{2^4}{2^2 \cdot 2! \cdot 5} - \frac{2^6}{2^3 \cdot 3! \cdot 7} + \frac{2^8}{2^4 \cdot 4! \cdot 9} = 0{,}61693216$$

и находим

$$\Phi(2) = \sqrt{\frac{2}{\pi}} \cdot 2 \cdot 0{,}616931216 = 0{,}984479744; \quad F(2) = \frac{1 + \Phi(2)}{2} = 0{,}992239892.$$

Относительная погрешность $\delta = 1{,}5\%$ велика.

Аппроксимация 5. Вычисляем $\Phi(2) = \sqrt{1 - \exp\left(-\dfrac{2 \cdot 2^2}{\pi}\right)} = \sqrt{1 - e^{-2{,}546479089}} = 0{,}960022359$.

Имеем $F(2) = \dfrac{1 + \Phi(2)}{2} = 0{,}98011179$.

Относительная погрешность $\delta = 0{,}28\%$ велика, так как аппроксимация применена для $z > 1{,}96$; $\Phi(z) > 0{,}95$, т. е. за пределами допустимого диапазона.

Аппроксимация 6. Вычисляем последовательно $1 + 0{,}1253 \cdot 2^2 = 1{,}5012$; $\left(\dfrac{2}{\pi} - 0{,}1253\right) \times 2^2 = 2{,}045279089$; $e^{-2{,}045279089} = 0{,}129344086$; $1 - \dfrac{0{,}129344086}{1{,}5012} = 0{,}913839537$.

Далее имеем $\Phi(2) = \sqrt{0{,}913839537} = 0{,}955949547$ и окончательно

$$F(2) = \frac{1 + \Phi(2)}{2} = 0{,}977974773.$$

Относительная погрешность $\delta = 0{,}07\%$.

Аппроксимация 7. Вычисляем последовательно $0{,}27893 \cdot \dfrac{2}{\sqrt{2}} = 0{,}393707156$; $0{,}23089 \times \dfrac{4}{2} = 0{,}460778$; $0{,}000972 \cdot \dfrac{2^3}{2\sqrt{2}} = 2{,}7492311 \cdot 10^{-3}$; $0{,}078108 \cdot \dfrac{2^4}{4} = 0{,}312432$.

Окончательно вычисляем

$$F(2) = 1 - 0{,}5(1 + 0{,}393707156 + 0{,}460778 + 2{,}7492311 \cdot 10^{-3} + 0{,}312432)^{-4} = \\ = 1 - 0{,}5 \cdot 2{,}169666387^{-4} = 0{,}977436931.$$

Относительная погрешность $\delta = 0{,}02\%$.

Аппроксимация 8. Вычисляем $\left(\dfrac{2+1{,}5774}{2{,}0637}\right)^{2{,}34} = 3{,}623055577$ и $e^{-3{,}623055577} = 0{,}026700965$.

Окончательно находим $F(2) = 1 - 0{,}852 \cdot 0{,}026706965 = 0{,}977250777$.

Относительная погрешность $\delta = 0{,}00007\%$.

Аппроксимация 18. Вычисляем $-0{,}717 \cdot 2 - 0{,}416 \cdot 2^2 = -3{,}098$; $e^{-3{,}098} = 0{,}04513939$. Находим $F(2) = 1 - 0{,}5 \cdot 0{,}04513939 = 0{,}97743005$.

Относительная погрешность $\delta = 0{,}018\%$.

Аппроксимация 20. Вычисляем $F(2) = \left\{1 + \exp\left[-0{,}0725 \cdot 2 \cdot \left(22 + 2^{1{,}96}\right)\right]\right\}^{-1} = 0{,}977115384$.

Относительная погрешность $\delta = 0{,}014\%$. Это одна из наиболее простых и точных аппроксимаций.

Задача 2. *Для случайной величины, имеющей стандартное нормальное распределение, найти значение, вероятность превышения которого равна* 0,05.

Другими словами, требуется найти значение u_p: $\mathbf{P}(x \leqslant u_p) = 0{,}95$, или $\mathbf{P}(x > u_p) = 0{,}05$. По определению u_p — верхняя 95%-я квантиль случайной величины. Воспользуемся аппроксимациями для стандартного нормального распределения. Табличное значение его квантили $u_{0,95} = 1{,}64485363$.

Аппроксимация 9. Так как эта аппроксимация применима только для значений $p \leqslant 0{,}01$, воспользуемся соотношением $u_p = -u_{1-p}$ и будем искать квантиль $u_{1-p} = u_{0,95}$, имея в виду, что $u_{0,95} = -u_{0,05}$.

Вычисляем

$$u_{0,05} = -\left[-2\ln(\sqrt{2\pi} \cdot 0{,}05)\right]^{\frac{1}{2}} = -\left[-2\ln 0{,}125331413\right]^{\frac{1}{2}} = -2{,}038035201.$$

Следовательно, $u_{0,95} = 2{,}038035201$.

Относительная погрешность $\delta = 23{,}95\%$ очень велика. Это объясняется тем, что аппроксимация применена вне допустимых пределов (должно быть $p \leqslant 0{,}01$).

Аппроксимация 10. Вычисляем $t = [-2\ln(1-0{,}95)]^{\frac{1}{2}} = 2{,}447746831$ и $\sum\limits_{i=0}^{2} c_i t^i = 2{,}515517 + 0{,}802853 \cdot 2{,}447746831 + 0{,}010328 \cdot 2{,}447746831^2 = 4{,}542577732$.

Далее $\sum\limits_{i=1}^{3} d_i t^i = 1{,}432788 \cdot 2{,}447746831 + 0{,}189269 \cdot 2{,}447746831^2 + 0{,}001308 \cdot 2{,}447746831^3 = 4{,}66028338$.

Вычисляем $\dfrac{4{,}542577732}{1 + 4{,}66028338} = 0{,}80253539$ и окончательно $u_{0,95} = 2{,}447746831 - 0{,}80253539 = 1{,}64521144$.

Относительная погрешность $\delta = 0{,}02\%$.

Аппроксимация 11. Вычисляем последовательно $\sqrt{2\pi}p = \sqrt{2\pi} \cdot 0{,}05 = 0{,}125331413$; $\sqrt{0{,}125331413} = 0{,}35402177$; $-2\ln 0{,}35402177 = 2{,}07679374$; $\sqrt{2\pi} \cdot 0{,}05 \cdot 2{,}076679374 = 0{,}260287495$; $-2 \cdot \ln 0{,}260287495 = 2{,}69193701$; $u_{0,05} = -\sqrt{2{,}69193701} = -1{,}64071235$.

Следовательно, $u_{0,95} = -u_{0,05} = 1{,}64071235$.

Относительная погрешность $\delta = 0{,}25\%$.

Аппроксимация 12. В нашем случае $\frac{1+p}{2} = 0{,}95$, откуда $p = 0{,}9$ и $u_{0,95} = \sqrt{-\frac{\pi}{2}\ln(1-0{,}9^2)} = 1{,}615137313$.

Относительная погрешность $\delta = 1{,}8\%$.

Аппроксимация 13. Имеем по аналогии $\frac{1+p}{2} = 0{,}95$ и $p = 0{,}9$. Далее вычисляем $-\ln\left(1-0{,}9^{1{,}898}\right) = 1{,}707888074$; $1{,}707888074^{\frac{1}{1{,}898}} = 1{,}325793782$. Окончательно имеем $u_{0,95} = \sqrt{\frac{\pi}{2}} \cdot 1{,}325793782 = 1{,}66163609$.

Относительная погрешность $\delta = 1\%$.

Аппроксимация 14. По аналогии $\frac{1+p}{2} = 0{,}95$ и $p = 0{,}9$. Далее вычисляем $-\ln(1-0{,}9^4) = 1{,}067404362$; $(1{,}067404362)^{\frac{3}{2}} = 1{,}102791627$; $-\ln\sqrt{1-0{,}9^2} = 0{,}830365603$; $\pi\left(\frac{1}{41}\cdot 1{,}102791627 + 0{,}830365603\right) = 2{,}693171017$.

Окончательно имеем $u_{0,95} = \sqrt{2{,}693171017} = 1{,}641088363$.

Относительная погрешность $\delta = 0{,}2\%$.

Аппроксимация 15. Вычисляем $0{,}95^{0,14} = 0{,}9922844661$; $(1-0{,}95)^{0,14} = 0{,}65743951$. Следовательно, $u_{0,95} = 4{,}91(0{,}992844661 - 0{,}65743951) = 1{,}646839291$.

Относительная погрешность $\delta = 0{,}12\%$.

Аппроксимация 16. Вычисляем $\ln\frac{1}{1-0,95} - 0{,}16 = 2,835732274$; $2{,}835732274^{0{,}4274} = 1{,}561238015$.

Окончательно имеем $u_{0,95} = 2{,}0637 \cdot 1{,}561238015 - 1{,}5774 = 1{,}644526892$.

Относительная погрешность $\delta = 0{,}02\%$.

Аппроксимация 17. Вычисляем $-\ln(1-0{,}95)^{\frac{5}{4}} = 3{,}744665342$; $-\ln\sqrt{2\,(1-0{,}95)} = 1{,}151292546$; $\frac{3{,}744665342}{23} + 1{,}151292546 = 1{,}314104083$.

Окончательно вычисляем $u_{0,95} = \sqrt{\pi \cdot 1{,}314104083} = 2{,}031841464$.

Относительная погрешность $\delta = 23{,}5\%$ велика, так как значение $p = 0{,}05$ находится вне области применения этой аппроксимации.

Аппроксимация 19. В нашем случае $\frac{1+p}{2} = 0{,}95$ и $p = 0{,}9$.

Вычисляем $4\cdot(-0{,}416)\cdot\ln(1-0{,}9) = 3{,}831501595$ и $\sqrt{0{,}717^2 + 3{,}831501595} = 2{,}084608019$. Окончательно имеем $u_{0,95} = \frac{0{,}717 - 2{,}084608019}{-2\cdot 0{,}416} = 1{,}643759638$.

Относительная погрешность $\delta = 0{,}066\%$.

Аппроксимация 21. Вычисляем $K = -\ln\left(\frac{1}{0{,}95} - 1\right) = 2{,}944438979$. Далее имеем

$$u_{0,95} = \frac{1{,}24 + 0{,}85\cdot 2{,}944438979^{0{,}657}}{1 + 0{,}0001\cdot 2{,}944438979^3 + \dfrac{2{,}38}{2{,}944438979}} = 1{,}6639025.$$

Относительная погрешность $\delta = 0{,}35\%$.

Задача 3. *Найти математические ожидания 3-й, 7-й и 9-й порядковых статистик в выборке объема $n = 15$ из стандартного нормального распределения.*

Вычислим необходимые вероятности

— для третьей порядковой статистики ($i = 3$) $p_3 = \frac{3-3/8}{15+1/4} = 0{,}172131148$;

— для седьмой порядковой статистики ($i = 7$) $p_7 = \frac{7-3/8}{15+1/4} = 0{,}434426230$;

— для девятой порядковой статистики ($i = 9$) $p_9 = \dfrac{9-3/8}{15+1/4} = 0{,}565573770$.

Далее воспользуемся простыми, но достаточно точными аппроксимациями 15 и 16. В соответствии с аппроксимацией 15 находим

$$z_3 = 4{,}91\left[0{,}172131148^{0,14} - (1-0{,}172131147)^{0,14}\right] = -0{,}9438796;$$
$$z_7 = 4{,}91\left[0{,}434426230^{0,14} - (1-0{,}434426230)^{0,14}\right] = -0{,}1643847;$$
$$z_9 = 4{,}91\left[0{,}565573770^{0,14} - (1-0{,}565573770)^{0,14}\right] = 0{,}1643847.$$

При использовании аппроксимации 16 получаем (для z_3 и z_7 используем соответственно формулы для $p = 1 - p_3$ и $p = 1 - p_7$ с заменой знака перед u_p, так как $u_p = 1 - u_{1-p}$):

$$z_3 = u_{0,172131148} = -2{,}0637\left(\ln\frac{1}{0{,}172131148} - 0{,}16\right)^{0,4274} + 1{,}5774 = -0{,}94508893;$$
$$z_7 = u_{0,434426230} = -2{,}0637\left(\ln\frac{1}{0{,}434426230} - 0{,}16\right)^{0,4274} + 1{,}5774 = -0{,}16577582;$$
$$z_9 = u_{0,565573770} = 2{,}0637\left(\ln\frac{1}{1-0{,}565573770} - 0{,}16\right)^{0,4274} - 1{,}5774 = 0{,}16577582.$$

Задача 4. *Долговечность прибора подчиняется нормальному распределению со средним $\mu = 100$ ч и стандартным отклонением $\sigma = 50$ ч. Вычислить вероятность того, что долговечность прибора будет больше 200 ч; меньше 50 ч; будет находиться в интервале от 70 ч до 120 ч. Вычислить значение долговечности прибора, вероятность превышения которого равна 0,01; 0,95. Вычислить математическое ожидание 2-й и 11-й порядковых статистик в выборке из 20 испытываемых приборов.*

Переходим к нормированной переменной $z = \dfrac{x-100}{50}$. При $x = 200$ ч имеем $z = \dfrac{200-100}{50} = 2$. Вероятность того, что долговечность прибора будет превышать 200 ч, равна (по определению)

$$\mathbf{P}(x > 200) = 1 - \mathbf{P}(x \leqslant 200) = 1 - F(z) = 1 - F(2).$$

Используя полученное в задаче 1 значение $F(2) = 0{,}97725$, получаем $1 - F(2) = 0{,}02275$. Следовательно, вероятность того, что долговечность прибора превысит 200 ч, равна 0,02275.

При $x = 50$ имеем $z = \dfrac{50-100}{50} = -1$. Вероятность того, что долговечность прибора не превысит 50 ч, равна $\mathbf{P}(x < 50) = F(z) = F(-1) = 1 - F(1)$. Используя простую аппроксимацию 18, вычисляем

$$F(1) = 1 - \frac{1}{2}\exp\left(-0{,}714\cdot 1 - 0{,}416\cdot 1\right) = 0{,}838967;$$
$$\mathbf{P}\,(x < 50) = 1 - 0{,}838967 = 0{,}16103.$$

Следовательно, вероятность того, что долговечность прибора будет меньше 50 ч, равна 0,16103.

Для вычисления вероятности попадания долговечности прибора в интервал (70; 120] ч определим $z_1 = \dfrac{70-100}{50} = -0{,}6$ и $z_2 = \dfrac{120-100}{50} = 0{,}4$.

Имеем $\mathbf{P}(70 < x \leqslant 120) = F(0{,}4) - F(-0{,}6) = F(0{,}4) + F(0{,}6) - 1$.

Воспользуемся аппроксимацией 8:

$$F(0{,}6) = 1 - 0{,}852\exp\left[-\left(\frac{0{,}6+1{,}5774}{2{,}0637}\right)^{2,34}\right] = 0{,}725795209;$$

$$F(0{,}4) = 1 - 0{,}852 \exp\left[-\left(\frac{0{,}4+1{,}5774}{2{,}0637}\right)^{2{,}34}\right] = 0{,}655128697.$$

Находим

$$\mathbf{P}(70 < x \leqslant 120) = 0{,}725795209 + 0{,}655286976 - 1 = 0{,}321082197.$$

Следовательно, вероятность того, что долговечность прибора будет находиться в интервале (70; 120] ч, равна 0,381.

Найдем теперь значение долговечности прибора, вероятность превышения которого равна 0,01. Это значение равно квантили распределения при $p = 0{,}99$. Воспользуемся аппроксимацией 16. При $p = 0,01$ имеем

$$u_{0{,}99} = 2{,}0637\left(\ln\frac{1}{1-0{,}99} - 0{,}16\right)^{0{,}4274} - 1{,}5774 = 2{,}3269978.$$

Это квантиль стандартного (нормированного) нормального распределения, а искомая величина является квантилью исходного ненормированного распределения. Переход к нему осуществляется по формуле

$$u'_{0{,}99} = \mu + u_{0{,}99}\sigma = 100 + 50\cdot 2{,}3266997 = 216{,}35.$$

Следовательно, с вероятностью 0,01 можно утверждать, что долговечность прибора превысит 216,35 ч.

Долговечность прибора, вероятность превышения которой 0,95, соответствует квантилю $u_{0{,}05} = -u_{0{,}95}$. В соответствии с аппроксимацией 16 получим

$$u_{0{,}95} = 2{,}0637\left(\ln\frac{1}{1-0{,}95} - 0{,}16\right)^{0{,}4247} - 1{,}5774 = 1{,}644527.$$

В нашем случае искомая стандартизованная величина квантили есть $u_{0{,}05} = -1{,}644527$, или для исходного ненормированного распределения

$$u'_{0{,}05} = \mu + u_{0{,}05}\sigma = 100 - 50\cdot 1{,}644527 = 17{,}77.$$

Таким образом, с вероятностью 0,95 долговечность прибора будет больше 17,8 ч.

Вычислим теперь математическое ожидание 2-й и 11-й порядковых статистик в выборке испытываемых приборов численностью $n = 20$. Имеем

$$\text{для}\quad i=2\quad p_2 = \frac{2-\frac{3}{8}}{20+\frac{1}{4}} = 0{,}0802469;\quad \text{для}\quad i=11\quad p_{11} = \frac{11-\frac{3}{8}}{20+\frac{1}{4}} = 0{,}52469138.$$

Тогда с помощью аппроксимации 16 имеем

$$z_2 = u_{0{,}0802469} = -2{,}0637\left(\ln\frac{1}{0{,}0802469} - 0{,}16\right)^{0{,}4274} + 1{,}5774 = -1{,}402743931,$$

или для исходного распределения

$$x_{(2)} = 100 - 50\cdot 1{,}402743931 = 29{,}86 \text{ ч.}$$

По аналогии для z_{11} имеем

$$z_{11} = u_{0{,}52469136} = 2{,}0637\left(\ln\frac{1}{1-0{,}52469136} - 0{,}16\right)^{0{,}4274} - 1{,}5774 = -0{,}062226575,$$

или для исходного распределения

$$x_{(11)} = 100 - 50\cdot 0{,}062226575 = 96{,}89 \text{ ч.}$$

Таким образом, при испытании 20 приборов с заданным распределением долговечности математические ожидания долговечности 2-го и 11-го приборов, ранжированных в порядке возрастания долговечности, будут равны соответственно 29,9 ч и 96,9 ч.

1.1.2. Равномерное распределение

Описание, применение. Находит широкое применение в непараметрической статистике. Равномерному распределению подчиняются случайные величины, имеющие одинаковую вероятность появления (например, погрешность измерений с округлением).

Свойства

Обозначение	$R(a, b)$
Параметры	a, b
Плотность	$f(x; a, b) = \begin{cases} (b-a)^{-1}, & a < x < b; \\ 0, & x < a; \quad x > b \end{cases}$
Функция распределения	$F(x; a, b) = \begin{cases} 0, & x \leqslant a; \\ \dfrac{x-a}{b-a}, & a < x < b; \\ 1, & x \geqslant b \end{cases}$
Среднее	$\mathbf{M}(x) = \dfrac{b+a}{2}$
Дисперсия	$\mathbf{D}(x) = \dfrac{(b-a)^2}{12}$
Коэффициент вариации	$v = \dfrac{1}{\sqrt{3}} \dfrac{b-a}{b+a}$
Коэффициент асимметрии	$\alpha_3 = 0$
Коэффициент эксцесса	$\alpha_4 = 1{,}8$
Медиана	$\mathbf{Me} = \dfrac{b+a}{2} = \mathbf{M}(x)$
Мода не определена	

Сумма n независимых равномерно распределенных случайных величин описывается нормальным распределением уже при $n \geqslant 5$. Функция распределения любой случайной величины $y - F(y)$ сама распределена равномерно на отрезке $[0, 1]$.

Задача 5. *Погрешность измерения прибора распределена равномерно на интервале* $[5; 10]$. *Найти вероятность того, что погрешность прибора не превышает* 7 *ед. Вычислить погрешность измерения, вероятность превышения которой равна* 0,95. *Вычислить вероятность того, что погрешность измерения будет находиться в интервале* $6 \div 8$ *ед.*

Имеем равномерно распределенную случайную величину с параметрами распределения $a = 5$ и $b = 10$.

Вероятность того, что погрешность не превысит $x = 7$ единиц, равна $F(7) = \dfrac{7-5}{10-5} = 0{,}4$. Значение случайной величины, вероятность превышения которого равна 0,95, находим из условия $F(x) = \dfrac{x-a}{b-a} = 1 - 0{,}95 = 0{,}05$, откуда $x_{0,95} = (10-5) \cdot 0{,}05 + 5 = 5{,}25$. Вероятность того, что погрешность измерения будет находиться в интервале $[6; 8]$, находим из условия

$$\mathbf{P}(6 \leqslant x \leqslant 8) = F(8) - F(6) = \frac{8-5}{10-5} - \frac{6-5}{10-5} = 0{,}4.$$

1.1.3. Логарифмически нормальное распределение

Описание, применение. Если случайная величина Y распределена нормально, то случайная величина $x = \ln Y$ подчинена логарифмически нормальному (или логнормальному) закону распределения. Часто используется для описания износовых отказов. У многих невосстанавливаемых электронных приборов (некоторые типы электронных ламп, полупроводниковые приборы) наработка на отказ распределена логарифмически нормально.

Свойства

Обозначение	$LN(\mu, \sigma)$
Параметры	μ, σ
Плотность	$f(x; \mu, \sigma) = \dfrac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\dfrac{(\ln x - \mu)^2}{2\sigma^2}\right\}, \quad x, \sigma > 0$
Функция распределения	$F(x; \mu, \sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} \int\limits_0^x \exp\left\{-\dfrac{(\ln t - \mu)^2}{2\sigma^2}\right\} dt; \quad x > 0$
Среднее	$\mathbf{M}(x) = \exp\left\{\mu + \dfrac{1}{2}\sigma^2\right\}$
Дисперсия	$\mathbf{D}(x) = \exp\{2\mu + \sigma^2\}\left(e^{\sigma^2} - 1\right)$
Коэффициент вариации	$v = \left(e^{\sigma^2} - 1\right)^{\frac{1}{2}}$
Коэффициент асимметрии	$\alpha_3 = \left(e^{\sigma^2} - 1\right)^{\frac{1}{2}}\left(e^{\sigma^2} + 2\right)$
Коэффициент эксцесса	$\alpha_4 = 3 + \left(e^{\sigma^2} - 1\right)\left(e^{3\sigma^2} + 3e^{2\sigma^2} + 6\right)$
Мода	$\mathbf{Mo} = \exp(\mu - \sigma^2)$
Медиана	$\mathbf{Me} = e^{\mu}$

Распределение имеет положительную асимметрию. Произведение независимых случайных величин, подчиняющихся логарифмически нормальному распределению, также распределено логарифмически нормально.

Логарифмически нормальное распределение иногда ошибочно принимается за экспоненциальное [47]. При вычислениях, связанных с логарифмически нормальным распределением, пользуются приемами для нормального распределения, заменяя при этом значение случайной величины ее логарифмом. Подробный анализ этого распределения приведен в [42].

Укажем приблизительные критерии проверки логнормальности распределения [43]. Распределение случайной величины близко к логнормальному, если

$$\frac{1}{n\sigma_{\ln x}} \sum |\ln x - \overline{\ln x}| = \sqrt{\frac{2}{\pi}} = 0{,}79788$$

или

$$\lg \mathbf{Me} = \overline{\ln x}, \quad \text{где} \quad \overline{\ln x} = \frac{1}{n}\sum_{i=1}^{n} \ln x_i \quad \text{и} \quad \sigma_{\ln x} = \frac{1}{n-1}\sum(\ln x_i - \overline{\ln x_i})^2.$$

Задача 6. *Предположим, что время безотказной работы элемента прибора — случайная величина, подчиняющаяся логарифмически нормальному закону распределения вероятностей с медианой, равной* 1000 *ч, и модой, равной* 400 *ч. Вычислить вероятность того, что элемент будет работать меньше* 2000 *ч.*

Имеем $\mathbf{Me} = e^{\mu} = 1000$, или $\mu = \ln 1000 = 6{,}908$; $\mathbf{Mo} = \exp(\mu - \sigma^2) = 400$, или $\mu - \sigma^2 = \ln 400 = 5{,}991$ и $\sigma^2 = 8{,}908 - 5{,}991 = 0{,}917$ $(\sigma = 0{,}958)$.

Воспользуемся нормированной случайной величиной $z = \dfrac{\ln x - \mu}{\sigma}$ и для $x = 2000$ имеем $z = \dfrac{\ln 2000 - 8{,}908}{0{,}958} = 0{,}723$. Применяя аппроксимацию 8 для нормального распределения, получим

$$\mathbf{P}(x < 2000) = \mathbf{P}(z < 0{,}723) = F(z) = 1 - 0{,}852\exp\left\{-\left(\frac{0{,}723 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\} =$$
$$= 1 - 0{,}852\exp(1 - 1{,}289278973) = 0{,}76553.$$

1.1.4. Экспоненциальное распределение

Описание, применение. Одно из наиболее часто встречающихся распределений в теории надежности и в теории массового обслуживания. Используется для описания внезапных отказов, когда износом изделия можно пренебречь. Наработка на отказ многих невосстанавливаемых изделий и наработка между соседними отказами у восстанавливаемых изделий в случае простейшего потока отказов подчинены экспоненциальному распределению. Наработка на отказ большой многокомпонентной системы может быть описана экспоненциальным распределением при любом распределении наработки на отказ компонентов системы.

Свойства

Параметр	b
Плотность	$f(x;b) = \dfrac{1}{b}\exp\left(-\dfrac{x}{b}\right), \quad x \geqslant 0$
Функция распределения	$F(x,b) = 1 - \exp\left(-\dfrac{x}{b}\right), \quad x \geqslant 0$
Среднее	$\mathbf{M}(x) = b$
Дисперсия	$\mathbf{D}(x) = b^2$
Коэффициент вариации	$v = 1$
Коэффициент асимметрии	$\alpha_3 = 2$
Коэффициент эксцесса	$\alpha_4 = 9$
Мода	$\mathbf{Mo} = 0$
Медиана	$\mathbf{Me} = b\ln 2 = 0{,}6931b$

Экспоненциальное распределение является частным случаем гамма-распределения (см. раздел 1.1.6) и распределения Вейбулла (см. раздел 1.1.5). Отличительная особенность экспоненциального распределения — постоянство интенсивности отказов $1/b = \mathrm{const}$ — в теории надежности интерпретируется как независимость вероятности отказа от наработки, что эквивалентно отсутствию износа.

Задача 7. *Наработка на отказ прибора распределена экспоненциально с интенсивностью отказов* $\lambda = 10^{-51}$ *ч*$^{-1}$. *Вычислить вероятность того, что наработка на отказ превысит* 1000 *ч. Найти вероятность того, что наработка на отказ будет находиться в интервале от* 1200 *до* 1500 *ч. Вычислить значение наработки, вероятность превышения которой* 0,8. *Определить, как изменится наработка прибора при уменьшении интенсивности отказов до* $\lambda = 0{,}5 \cdot 10^{-5}$ *ч*$^{-1}$.

Вероятность того, что наработка на отказ превысит 1000 ч, равна

$$\mathbf{P}(x > 1000) = 1 - F(1000) = 1 - 1 + e^{-1000\cdot 10^{-5}} = e^{-10^{-2}} = 0{,}99.$$

Вероятность того, что наработка будет находиться в интервале от 1200 ч до 1500 ч, определяем по формуле

$$\mathbf{P}(1200 < x < 1500) = F(1500) - F(1200) = 1 - e^{-1500\cdot 10^{-5}} - 1 + e^{-1200\cdot 10^{-5}} = \\ = e^{-0{,}012} - e^{-0{,}015} = 0{,}988071712 - 0{,}985111939 = 0{,}0029598.$$

Наработку y, вероятность превышения которой равна 0,8, находим из соотношения $\mathbf{P}(x > y) = 1 - F(y) = 0{,}8$. Отсюда имеем $1 - 1 + e^{-y\cdot 10^{-5}} = 0{,}8$, или $y\cdot 10^{-5} = 0{,}22314355$. Окончательно $y = 2{,}23 \cdot 10^4$ ч.

При снижении интенсивности отказов до $\lambda = 0{,}5\cdot 10^{-5}$ имеем $y\cdot 0{,}5\cdot 10^{-5} = 0{,}22314355$, или $y = 4{,}46 \cdot 10^4$ ч.

1.1.5. Распределение Вейбулла

Описание, применение. Этому распределению подчиняется наработка на отказ многих невосстанавливаемых электронных приборов (электронные лампы, полупроводниковые приборы, некоторые приборы СВЧ). Характеризуется разнообразием форм кривых распределения.

Свойства

Обозначение $W(\alpha, \beta)$

Параметры α, β

Плотность $f(x;\alpha,\beta) = \dfrac{\beta}{\alpha^\beta}\, x^{\beta-1} \exp\left[-\left(\dfrac{x}{\alpha}\right)^\beta\right], \quad x \geqslant 0, \quad \alpha,\beta > 0$

Функция распределения $F(x;\alpha,\beta) = 1 - \exp\left[-\left(\dfrac{x}{\alpha}\right)^\beta\right], \quad x \geqslant 0, \quad \alpha,\beta > 0$

Среднее $\mathbf{M}(x) = \alpha\Gamma\left(1 + \dfrac{1}{\beta}\right)$

Дисперсия $\mathbf{D}(x) = \alpha^2\left[\Gamma\left(1 + \dfrac{2}{\beta}\right) - \left\{\Gamma\left(1 + \dfrac{1}{\beta}\right)\right\}^2\right]$

Коэффициент вариации $v = \left[\dfrac{\Gamma\left(1 + \frac{2}{\beta}\right)}{\left\{\Gamma\left(1 + \frac{1}{\beta}\right)\right\}^2} - 1\right]^{\frac{1}{2}}$

Коэффициент асимметрии

$$\alpha_3 = \frac{\Gamma\left(1 + \frac{3}{\beta}\right) - 3\Gamma\left(1 + \frac{2}{\beta}\right)\Gamma\left(1 + \frac{1}{\beta}\right) + 2\left\{\Gamma\left(1 + \frac{1}{\beta}\right)\right\}^3}{\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left\{\Gamma\left(1 + \frac{1}{\beta}\right)\right\}^2\right]^2}$$

Коэффициент эксцесса

$$\alpha_4 = \frac{\Gamma\left(1 + \frac{4}{\beta}\right) - 4\Gamma\left(1 + \frac{3}{\beta}\right)\Gamma\left(1 + \frac{1}{\beta}\right) + 6\Gamma\left(1 + \frac{2}{\beta}\right)\left\{\Gamma\left(1 + \frac{1}{\beta}\right)\right\}^2 - 3\left\{\Gamma\left(1 + \frac{1}{\beta}\right)\right\}^4}{\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left\{\Gamma\left(1 + \frac{1}{\beta}\right)\right\}^2\right]^2}$$

Мода $\mathbf{Mo} = \alpha\left(1 - \dfrac{1}{\beta}\right)^{\frac{1}{\beta}}, \quad \beta > 1$

При $\beta = 1$ распределение Вейбулла переходит в экспоненциальное (см. раздел 1.1.4) с параметром α, а при $\beta = 2$ — в распределение Рэлея (см. раздел 1.1.15). Вычисление моментов распределения Вейбулла производится по таблицам гамма-функции. Таблицы функции и моментов распределения Вейбулла приведены в [44].

Задача 8. *Наработка прибора подчиняется распределению Вейбулла с параметрами* $\alpha = 2$ *и* $\beta = 3$. *Вычислить моду распределения и вероятность нахождения наработки прибора в интервале между* 5 *и* 6.

Находим моду распределения $\mathbf{Mo} = \alpha\left(1 - \frac{1}{\beta}\right)^{\frac{1}{\beta}} = 2\left(1 - \frac{1}{3}\right)^{\frac{1}{3}} = 1{,}74716.$

Далее

$$\mathbf{P}(5 \leqslant x \leqslant 6) = F(6) - F(5) = 1 - \exp\left\{-\left(\frac{6}{2}\right)^3\right\} - 1 + \exp\left\{-\left(\frac{5}{2}\right)^3\right\} =$$

$$= \exp\left\{-\left(\frac{5}{2}\right)^3\right\} - \exp\left\{-\left(\frac{6}{2}\right)^3\right\} = 1{,}6373771 \cdot 10^{-7} - 1{,}8795288 \cdot 10^{-12} = 1{,}637 \cdot 10^{-7}.$$

1.1.6. Гамма-распределение

Описание, применение. Широко используется в теории надежности и в теории массового обслуживания. Наработка между несмежными отказами подчиняется гамма-распределению. Этому распределению (с параметром $\alpha = r - 1$) подчиняется сумма r независимых случайных величин, каждая из которых имеет экспоненциальное распределение. Если наработка на отказ невосстанавливаемого прибора распределена экспоненциально, то, при испытаниях на безотказность с заменой отказавших приборов, момент r-го отказа подчиняется гамма-распределению с параметром $\alpha = r - 1$. Гамма-распределение с целочисленным значением параметра называется также распределением Эрланга.

Свойства

Обозначение $\gamma(\alpha, \beta)$

Параметры α, β

Плотность

$$f(x; \alpha, \beta) = \frac{1}{\alpha!\beta^{\alpha+1}} x^\alpha \exp\left(-\frac{x}{\beta}\right), \quad x > 0, \quad \beta > 0, \quad \alpha > 1$$

Функция распределения

$$F(x; \alpha, \beta) = \int\limits_0^x \frac{1}{\alpha!\beta^{\alpha+1}} x^\alpha \exp\left(-\frac{x}{\beta}\right) dx = \frac{1}{\alpha}\Gamma_{\frac{x}{\beta}}(\alpha + 1) = 1 - \sum_{i=0}^{\alpha} e^{-\frac{x}{\beta}} \left(\frac{x}{\beta}\right)^i \frac{1}{i!},$$

где $\Gamma_{\frac{x}{\beta}}(\alpha + 1)$ — неполная гамма-функция, значения которой приведены, например, в [45–47]

Среднее $\mathbf{M}(x) = \beta(\alpha + 1)$

Дисперсия $\mathbf{D}(x) = \beta^2(\alpha + 1)$

Коэффициент вариации $v = (\alpha + 1)^{-\frac{1}{2}}$

Коэффициент асимметрии $\alpha_3 = 2(\alpha + 1)^{-\frac{1}{2}}$

Коэффициент эксцесса $\alpha_4 = 3 + 6(\alpha + 1)^{-1}$

Мода $\mathbf{Mo} = \alpha\beta$

При $\alpha = 0$ гамма-распределение переходит в экспоненциальное (см. раздел 1.1.4). При использовании в теории надежности интенсивность отказов убывает при $\alpha < 0$, постоянна при $\alpha = 0$ и возрастает при $\alpha > 0$, что позволяет использовать гамма-распределение при моделировании различных ситуаций, возникающих в процессе анализа надежности изделий.

Гамма-распределение обладает свойством аддитивности, т. е. сумма независимых величин, подчиняющихся гамма-распределению с параметрами β и α_i, имеет также гамма-распределение с параметрами β и $\sum \alpha_i$. Подробно гамма-распределение исследовано в [48, 49].

При $(\alpha + 1)$ полуцелом (т. е. когда $2(\alpha + 1)$ — целое число) гамма-распределение можно рассматривать как частный случай распределения χ^2 (см. раздел 1.1.8) с $2(\alpha + 1)$ степенями свободы. Поэтому для расчетов, связанных с гамма-распределением, могут быть использованы таблицы и аппроксимации распределения χ^2. Распределению χ^2 с $f = 2(\alpha + 1)$ степенями свободы подчинена случайная величина $\gamma = \dfrac{2x}{\beta}$, т. е. квантиль случайной величины γ может быть вычислена как $\gamma_p = \dfrac{\beta}{2}\chi_p^2\big[2(\alpha + 1)\big]$, где $\chi_p^2\big[2(\alpha + 1)\big]$ — p-квантиль χ^2-распределения с $f = 2(\alpha + 1)$ степенями свободы.

Задача 9. *Испытываются четыре прибора, интенсивность отказов которых известна и равна $\lambda = 10^{-5}$ ч$^{-1}$. Вычислить вероятность того, что суммарная наработка приборов не превысит 300 000 ч.*

В нашем случае имеет место гамма-распределение с параметрами $\alpha = 4 - 1 = 3$ и $\beta = \dfrac{1}{\lambda} = 10^5$. Тогда искомая вероятность равна

$$\mathbf{P}(x < 300\,000) = F(300\,000) = 1 - e^{-3}\left(1 + 3 + 3^2 \cdot \frac{1}{2} + 3^3 \cdot \frac{1}{6}\right) = 0{,}352768.$$

1.1.7. Бета-распределение

Описание, применение. Часто используется в математической статистике, так как через бета-распределение могут быть выражены практически все применяемые распределения вероятностей, в том числе и дискретные. Доля дефектных изделий в партии подчиняется бета-распределению. Особенно велико значение бета-распределения в непараметрической статистике (т. е. при решении задач, не требующих знания закона распределения вероятностей случайной величины).

Свойства

Обозначение $\quad \mathrm{B}(\alpha, \beta)$

Параметры $\quad \alpha, \beta$

Плотность

$$f(x; \alpha, \beta) = \frac{(\alpha + \beta + 1)!}{\alpha!\,\beta!}\,x^{\alpha}(1 - x)^{\beta}, \quad 0 < x < 1; \quad \alpha, \beta \geqslant -1$$

Функция распределения

$$I_x(\alpha, \beta) = \int_0^x \frac{(\alpha + \beta + 1)!}{\alpha!\,\beta!}\,x^{\alpha}(1 - x)^{\beta}dx = \frac{(\alpha + \beta + 1)!}{\alpha!\,\beta!}\,\mathrm{B}_x(\alpha + 1, \beta + 1),$$

где $\mathrm{B}_x(\alpha + 1, \beta + 1) = \int_0^x x^{\alpha}(1 - x)^{\beta}dx$ — неполная бета-функция

Среднее $\mathbf{M}(x) = \dfrac{\alpha+1}{\alpha+\beta+2}$

Дисперсия $\mathbf{D}(x) = \dfrac{(\alpha+1)(\beta+1)}{(\alpha+\beta+2)^2(\alpha+\beta+3)}$

Коэффициент вариации $v = \left\{\dfrac{\beta+1}{(\alpha+1)(\alpha+\beta+3)}\right\}^{\frac{1}{2}}$

Коэффициент асимметрии $\alpha_3 = \dfrac{2(\beta-\alpha)}{\alpha+\beta+4}\left[\dfrac{\alpha+\beta+3}{(\alpha+1)(\beta+1)}\right]^{\frac{1}{2}}$

Коэффициент эксцесса

$$\alpha_4 = \frac{3(\alpha+\beta+2)(\alpha+\beta+3)}{(\alpha+\beta+4)(\alpha+1)(\beta+1)}\left[\frac{(\alpha+2)(-\alpha+2\beta+1)}{\alpha+\beta+5} + \frac{(\alpha+1)(\alpha-\beta)}{\alpha+\beta+2}\right]$$

Мода $\mathbf{Mo} = \dfrac{\alpha}{\alpha+\beta}$

Наиболее компактно функция бета-распределения может быть записана с использованием бета-функции Эйлера

$$\mathrm{B}(\alpha,\beta) = \int\limits_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Тогда

$$I_x(\alpha,\beta) = \frac{1}{B(\alpha,\beta)}B_x(\alpha+1,\beta+1).$$

Распределение симметрично при $\alpha=\beta$, $\alpha_3>0$ при $\alpha<\beta$ и $\alpha_3<0$ при $\alpha>\beta$. Широкое применение бета-распределения вызвано чрезвычайным разнообразием кривых распределения, порождаемых функцией бета-распределения при различных сочетаниях его параметров. При $\alpha=\beta=0$ бета-распределение превращается в равномерное, а при $\alpha=\beta=-1/2$ — в распределение арксинуса.Через бета-распределение могут быть выражены функция распределения Фишера (F-распределение, см. раздел 1.1.10)

$$F(x;f_1,f_2) = I_{\frac{f_1 x}{f_2+f_1 x}}\left(\frac{f_1}{2};\frac{f_2}{2}\right)$$

и функция биномиального распределения (см. раздел 1.2.1)

$$\sum_{i=0}^{k-1} C_n^i p^i(1-p)^{n-i} = 1 - I_p(k, n-k+1).$$

Из приведенных соотношений следует связь между бета- и F-распределениями:

$$I_x(f_1,f_2) = F\left(\frac{f_2}{f_1}\frac{x}{1-x};2f_1,2f_2\right).$$

Следовательно, случайная величина $\mathrm{B}(f_1,f_2) = \dfrac{f_1}{f_1+f_2F(f_1,f_2)}$ имеет бета-распределение, или (что эквивалентно) величина

$$F(2f_1,2f_2) = \frac{f_2B(f_1,f_2)}{f_1B(f_1,f_2)}$$

имеет F-распределение с $2f_1$ и $2f_2$ степенями свободы. Отметим также, что

$$I_x(\alpha, \beta) = 1 - (1-x)^{\alpha+\beta-1} \sum_{i=0}^{\alpha-1} C^i_{\alpha+\beta-1} \left(\frac{x}{1-x}\right)^i.$$

Для расчетов используются таблицы неполной бета-функции [50]; таблицы функции и квантили бета-распределения приведены в [51].

Если x_1 и x_2 — случайные величины, подчиненные гамма-распределению (см. раздел 1.1.6) с f_1 и f_2 степенями свободы, то случайная величина $B = x_1/(x_1+x_2)$ имеет бета-распределение с параметрами f_1 и f_2. Поскольку $I_x(\alpha,\beta) = 1 - I_{1-x}(\beta,\alpha)$, таблицы бета-распределения составлены для $0 < \alpha \leqslant \beta$. Однако разнообразие задач прикладного математико-статистического анализа в настоящее время не удовлетворяется существующими таблицами бета-распределения. Поэтому на практике применяются различные приближения, позволяющие вычислить бета-распределение с помощью таблиц или аппроксимаций нормального распределения. Приведем некоторые аппроксимации.

Аппроксимация Кэдуэлла [52] (при $\alpha = \beta$)

$$I_x(\alpha, \alpha) = F(y) + \frac{2(\alpha-1)(2\alpha+1)}{(4\alpha-1)^4}\varphi(y),$$

где y — решение уравнения

$$x = \frac{1}{2} + \left(\frac{\pi}{3}\right)^{\frac{1}{2}} \left[F\left\{\left(\frac{3}{4\alpha-1}\right)^{\frac{1}{2}} y\right\} - \frac{1}{2}\right]$$

и $F(y)$ — функция стандартного нормального распределения.

Более удобна эквивалентная формула

$$y = \left(\frac{3}{4\alpha-1}\right)^{-\frac{1}{2}} u_p,$$

где u_p — p-квантиль стандартного нормального распределения (см. раздел 1.1.1) и

$$p = \frac{x - \frac{1}{2}}{\left(\frac{\pi}{3}\right)^{\frac{1}{2}}} + \frac{1}{2}.$$

Наиболее употребляемые значения функции $\varphi(y)$ приведены в табл. 2 [50].

Таблица 2

Значения $\varphi(y)$

y	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,000	0,040	0,079	0,118	0,156	0,191	0,226	0,258	0,288	0,315
1	0,339	0,360	0,377	0,391	0,400	0,406	0,406	0,403	0,395	0,383
2	0,367	0,348	0,326	0,302	0,272	0,249	0,222	0,196	0,171	0,147
3	0,125	0,105	0,087	0,071	0,057	0,046	0,036	0,028	0,022	0,017

При $\alpha \geqslant 5$ погрешность не более $1 \cdot 10^{-5}$, при $\alpha \geqslant 4$ $I_x(\alpha,\alpha) = F(y)$ дает погрешность не более 0,00045.

Приведем еще одну полезную формулу:

$$I_x(\alpha,\alpha) = 2I_{x'}(\alpha,\alpha), \quad \text{где} \quad x' = \frac{1}{2}\left[1 - (1-x)^{\frac{1}{2}}\right].$$

Аппроксимация Уайза [43]. При $\alpha \geqslant \beta$ позволяет выразить бета-распределение через χ^2-распределение:

$$I_x(\alpha,\beta) = \mathbf{P}(\chi^2_{2\beta} > 2y) + \frac{y^\beta e^{-y}}{\Gamma(\beta-1)} \times$$
$$\times \left\{ \frac{1}{24N^2}(\beta+y+1) - \frac{1}{5760N^4}\left[(\beta-3)(\beta-2)(5\beta+7)(\beta+y+1) - (5\beta-7)(\beta+3+y)y^2\right] \right\},$$

где $N = \alpha + \dfrac{\beta}{2} - \dfrac{1}{2}$; $y = -N\ln x$ и $\chi^2_{2\beta}$ имеет χ^2-распределение с 2β степенями свободы (см. раздел 1.1.18).

Еще одна аппроксимация Уайза [54]

$$I_x(\alpha,\beta) = F(z),$$

если

$$z = \frac{\alpha}{|\beta - 0{,}5 - n(1-x)|}\left\{ \frac{2}{1+\dfrac{1}{6n}}\left[(\beta-0{,}5)\ln\frac{\beta-0{,}5}{n(1-x)} + (\alpha-0{,}5)\ln\frac{\alpha-0{,}5}{nx}\right]\right\}^{\frac{1}{2}},$$

где $n = \alpha+\beta-1$.

Приведем эквивалентную формулу

$$z_1 = d_1\frac{\left\{1 + zg\left(\dfrac{\beta-0{,}5}{n(1-z)}\right) + (1-z)g\left(\dfrac{\alpha-0{,}5}{nz}\right)\right\}^{\frac{1}{2}}}{\left(n+\dfrac{1}{6}\right)z(1-z)},$$

где $d_1 = \beta - 0{,}5 + \dfrac{1}{6} - \left(n+\dfrac{1}{3}\right)(1-x)$ и $g(x) = \dfrac{1-x^2+2x\ln x}{(1-x)^2}$.

Лучший результат дает аппроксимация для z_2, получаемая заменой в формуле для z_1 параметра d_1 на d_2, где

$$d_2 = d_1 + \frac{1}{50}\left\{\frac{x}{\beta} - \frac{1-x}{\alpha} + \frac{x-0{,}5}{\alpha+\beta}\right\}.$$

Погрешность этой аппроксимации $< 0{,}001$ при $\alpha,\beta \geqslant 2{,}0$ и $< 0{,}1$ при $\alpha,\beta \geqslant 1{,}0$.

Аппроксимация Кемпа–Полсона [55]. Случайная величина y имеет стандартное нормальное распределение $N(0,1)$, если

$$y = 3\frac{(\beta x)^{\frac{1}{3}}\left(1-\dfrac{1}{9\beta}\right) - \left[\alpha(1-x)^{\frac{1}{3}}\right]\left(1-\dfrac{1}{9\alpha}\right)}{\left\{\left[\dfrac{(1-x)^2}{\alpha}\right]^{\frac{1}{3}} + \left(\dfrac{x^2}{\beta}\right)^{\frac{1}{3}}\right\}^{\frac{1}{2}}},$$

где x — случайная величина, имеющая бета-распределение с параметрами α и β.

Задача 10. *Вычислить значение функции бета-распределения с параметрами $\alpha = 4$ и $\beta = 3$ в точке $x = 0{,}6$.*

Используем формулу для прямого точного вычисления

$$I_x(\alpha,\beta) = 1 - (1-x)^{\alpha+\beta-1}\sum_{i=0}^{\alpha-1} C^i_{\alpha+\beta-1}\left(\frac{x}{1-x}\right)^i =$$
$$= 1 - 1(1-0{,}6)^{3+4-1}\sum_{i=0}^{3} C^i_{4+3-1}\left(\frac{0{,}6}{1-0{,}6}\right)^i = 1 - 0{,}4\sum_{i=0}^{3} C^i_6(1{,}5)^i =$$
$$= 1 - 4{,}096\cdot 10^3\left(C^0_6 + C^1_6\cdot 1{,}5 + C^2_6\cdot 2{,}25 + C^3_6\cdot 3{,}75\right) = 0{,}54432.$$

Рассмотрим случай $\alpha = \beta = 3$ и используем аппроксимацию Кэдуэлла. Будем искать $I_{0,6}(3,3)$. Находим $p = \dfrac{0{,}6-0{,}5}{\sqrt{\pi/3}} + 0{,}5 = 0{,}5977205$. Для вычисления u_p используем аппроксимацию 16 из раздела 1.1.1

$$u_p = 2{,}0637\left(\ln\frac{1}{1-0{,}5977205} - 0{,}16\right)^{0{,}4274} - 1{,}5574 = 0{,}26816878.$$

Тогда $y = \left(\dfrac{3}{4\cdot 3-1}\right)^{-\frac{1}{2}} u_p = 0{,}513504118.$

Для вычисления $F(y)$ воспользуемся аппроксимацией 18 из раздела 1.1.1

$$F(y) = 1 - \frac{1}{2}\exp\left(-0{,}717y - 0{,}416y^2\right) = 0{,}68995.$$

Из табл. 2 имеем $\varphi(y) \approx 0{,}192$.

Окончательно получаем

$$I_{0,6}(3{,}3) = 0{,}68995 + \frac{2\cdot(3-1)\cdot(2\cdot 3+1)}{(4\cdot 3-1)^4}\cdot 0{,}192 = 0{,}6903.$$

Задача 11. *Вычислить значение $I_{0,6}(5,4)$ с помощью аппроксимации Уайза.*

Имеем $x = 0{,}6$; $\alpha = 5$; $\beta = 4$. Находим $n = 5 + 4 - 1 = 8$.

Далее имеем $\alpha = 4 - 0{,}5 + \dfrac{1}{6} - \left(1 + \dfrac{1}{3}\right)\cdot 0{,}4 = \dfrac{1}{3}$ и

$$\frac{\beta - 0{,}5}{n\cdot(1-x)} = \frac{3{,}5}{8\cdot 0{,}4} = 1{,}09375; \qquad \frac{\alpha-0{,}5}{n\cdot x} = \frac{\alpha - 0{,}5}{8\cdot 0{,}6} = 0{,}9375.$$

Далее имеем

$$g(1{,}09375) = \frac{1 - 1{,}09375^2 + 2\cdot 1{,}09375\ln 1{,}09375}{(1-1{,}09375)^2} =$$
$$= (1 - 1{,}19628906 + 0{,}196026597)\cdot 113{,}7777 = -0{,}029862726;$$

$$g(0{,}9375) = \frac{1 - 0{,}9375^2 + 2\cdot 0{,}9375\ln 0{,}9375}{(1-0{,}9375)^2} = 0{,}021509854.$$

Вычисляем

$$z_1 = \frac{1}{3}\cdot\frac{[1 + 0{,}6\cdot(-0{,}029862726) + 0{,}4\cdot 0{,}021509854]^{\frac{1}{2}}}{\left(8 + \frac{1}{6}\right)\cdot 0{,}4\cdot 0{,}6} = 0{,}169274.$$

Окончательно имеем

$$F(z) = 1 - \frac{1}{2}\exp(-0{,}717\cdot 0{,}169274 - 0{,}416\cdot 0{,}159274^2) = 0{,}562.$$

Точное значение $I_{0,6}(5{,}4) = 0{,}6$. Точность аппроксимации может быть повышена применением более точных аппроксимаций входящих величин (функции и квантилей стандартного нормального распределения).

Задача 12. *Вычислить* $I_{0,6}(5,4)$ *с помощью аппроксимации Кемпа–Полсона.*

Вычисляем

$$y = 3 \cdot \frac{(4\cdot 0{,}6)^{\frac{1}{3}} \cdot \left(1 - \frac{1}{9\cdot 4}\right) - (5\cdot 0{,}4)^{\frac{1}{3}} \cdot \left(1 - \frac{1}{9\cdot 5}\right)}{\left\{\left(\frac{0{,}4^2}{5}\right)^{\frac{1}{3}} + \left(\frac{0{,}6^2}{4}\right)^{\frac{1}{3}}\right\}^{\frac{1}{2}}} = 0{,}239151678.$$

Используя аппроксимацию 15 (см. раздел 1.1.1) для u_p, имеем

$$u_p = -4{,}91 \cdot \left[0{,}239151678^{0,14} - (1 - 0{,}239151678)^{0,14}\right] = 0{,}7068,$$

что не очень близко к точному значению $I_{0,6}(5,4)$.

1.1.8. Распределение χ^2 (распределение Пирсона)

Описание, применение. Распределение открыто и изучено Пирсоном в 1900 г. Если $x_1, \ldots, x_f$ — независимые случайные величины, имеющие стандартное нормальное распределение, то сумма их квадратов $\sum_{i=1}^{f} x_i^2$ подчиняется χ^2-распределению. Распределение хи-квадрат широко используется в прикладных задачах математической статистики. С его помощью проверяются гипотезы относительно значений дисперсий, проверяется согласие экспериментальных данных с теоретическими законами распределения. Распределение хи-квадрат широко применяется в непараметрической статистике, являясь предельным для многих выборочных статистик.

Свойства

Обозначение	$\chi^2(f)$
Параметр	f
Плотность	$\varphi(\chi^2; f) = \dfrac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{f}{2}\right)} (\chi^2)^{\frac{f-2}{2}} \exp\left\{-\dfrac{\chi^2}{2}\right\}, \quad \chi^2 \geqslant 0$
Функция распределения	$F_f(x) = \mathbf{P}\{\chi^2(f), x\} = \dfrac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{f}{2}\right)} \int_0^x y^{\frac{f}{2}-1} \exp\left\{-\dfrac{y}{2}\right\} dy,$ $x > 0$
Среднее	$\mathbf{M}\left[\chi^2(f)\right] = f$
Дисперсия	$\mathbf{D}[\chi^2(f)] = 2f$
Коэффициент вариации	$v = \left(\dfrac{2}{f}\right)^{\frac{1}{2}}$
Коэффициент асимметрии	$\alpha_3 = 2\left(\dfrac{2}{f}\right)^{\frac{1}{2}}$

Коэффициент эксцесса $\alpha_4 = 3 + \dfrac{12}{f}$

Мода $\mathbf{Mo} = f - 2, \quad f \geqslant 2$

Распределение имеет один параметр f — число степеней свободы, определяемый количеством независимых случайных величин, сумма квадратов которых составляет χ^2. Плотность распределения χ^2 асимметрична, унимодальна (т. е. имеет единственную моду) и с ростом числа степеней свободы f становится более пологой и симметричной.

χ^2-распределение обладает свойством аддитивности, т. е. сумма двух независимых величин $\chi^2(f_1)$ и $\chi^2(f_2)$ имеет распределение $\chi^2(f_1 + f_2)$.

Таблицы χ^2-распределения имеются во многих руководствах и сборниках таблиц [23–25, 29, 44, 56, 57]. Предложены номограммы для расчетов χ^2-распределения [58].

Большинство аппроксимаций квантилей χ^2-распределения основано на преобразовании исходной случайной величины в величину, имеющую распределение, близкое к стандартному нормальному $N(0,1)$.

Рассмотрим известные аппроксимации. Везде будем рассматривать верхнюю p-квантиль, т. е. величину χ_p^2, удовлетворяющую соотношению $p = 1 - \mathbf{P}\big(\chi^2 < \chi_p^2\big)$.

Напомним, что u_p — верхняя p-квантиль стандартного нормального распределения (т. е. $p = 1 - \mathbf{P}(x < u_p)$).

Аппроксимация 1. При $f > 200$ $\chi_p^2(f) = f + \sqrt{2f}\, u_p$.

Аппроксимация 2 (*Фишера*) [59, 60]. При $f > 100$ $\chi_p^2(f) = \dfrac{1}{2}\big(u_p + \sqrt{2f-1}\big)^2$.

Аппроксимация 3 (*Вилсона–Хилферти*) [61]:

$$\chi_p^2(f) = f\left[1 - \frac{2}{9f} + u_p\sqrt{\frac{2}{9f}}\right]^3, \qquad f \geqslant 30.$$

Аппроксимация 4 [62–64]. В [62] показано, что величина $(\chi^2)^{\frac{1}{4}}$ близка к нормальному распределению со средним $\mu = (f - 1/2)^{\frac{1}{4}}$ и дисперсией $\sigma^2 = (8f)^{-1}$. По сравнению с аппроксимацией 3, эта аппроксимация предпочтительнее для малых значений f и уступает ей при больших значениях f.

В [63, 64] показано, что для $(\chi^2)^{\frac{1}{\lambda}}$ оптимальное значение λ различно для различных f и равно:

f	1	2	3	4	6
λ	0,2084	0,2654	0,2887	0,3006	0,3124

Аппроксимация 5 (*Голдштейна*) [65]:

$$\chi_p^2(f) = f\left\{1 - \frac{2}{9f} + \frac{4x^4 + 16x^2 - 28}{1215f^2} + \frac{8x^6 + 720x^4 - 3216x^2 + 2904}{229635f^3} + \right.$$

$$\left. + \left(\frac{2}{f}\right)^{\frac{1}{2}} \left(\frac{\frac{x}{3} + (-x^3 + 3x)}{162f} - \frac{3x^5 + 40x^3 + 45x}{5832f^2} + \frac{301x^7 - 1519x^5 - 3269x^3 - 79349x}{7873200f^3}\right)\right\},$$

или в форме

$$\chi_p^2(f) = f\left[\sum_{i=0}^{6} f^{-\frac{f}{2}} x^i (a_i + b_i f^{-1} + c_i f^{-2})\right]^3,$$

где $x = u_p$ и

$a_0 = 1{,}0000886$; $b_0 = -0{,}2237368$; $c_0 = -0{,}01513904$;
$a_1 = 0{,}4713941$; $b_1 = 0{,}02607083$; $c_1 = -0{,}008986007$;
$a_2 = 0{,}0001348028$; $b_2 = 0{,}01128186$; $c_2 = 0{,}02277679$;
$a_3 = -0{,}008553069$; $b_3 = -0{,}01153761$; $c_3 = -0{,}01323293$;
$a_4 = 0{,}00312558$; $b_4 = 0{,}005169654$; $c_4 = -0{,}006950356$;
$a_5 = -0{,}0008426812$; $b_5 = 0{,}00253001$; $c_5 = 0{,}001060438$;
$a_6 = 0{,}00009780499$; $b_6 = -0{,}001450117$; $c_6 = 0{,}001565326$.

Аппроксимация 6 [66]. При $f > 10$

$$\chi_p^2(f) = f\left[1 - \frac{2}{9f} + (u_p - h_f)\sqrt{\frac{2}{9f}}\right]^3,$$

где

$$h_f = -\frac{2}{27f}\left(\frac{2\sqrt{2}(x^2-1)}{3\sqrt{f}} - \frac{x^3-3x}{4}\right), \quad x = u_p,$$

или в модифицированной форме

$$h'_f = \frac{(9f+16)(x^3-3x) - 24(x^2-1)\sqrt{2f}}{486f^2}.$$

Аппроксимация 7 (*Хэлдена*) [67]

$$\chi_p^2(f) = \frac{1}{13}\left\{12f\left[\frac{5x}{6\left(1-\frac{1}{18f}\right)12f} + 1 - \frac{5}{18f}\left(1+\frac{7}{48f}\right)\right]^{\frac{13}{5}} + f\right\}, \quad x = u_p.$$

Аппроксимация 8 (*Корниша–Фишера*) [68]. Упрощенная ($x = u_p$):

$$\chi_p^2(f) = f + u_p\sqrt{2f} + \frac{2}{3}\left(u_p^2 - 1\right) + \frac{1}{9\sqrt{2f}}\left(u_p^2 - 7u_p\right),$$

или

$$\chi_p^2(f) = f + G_1(x)\sqrt{f} + G_2(x) + G_3(x)\frac{1}{\sqrt{f}} + G_4(x)\frac{1}{f} + G_5(x)\frac{1}{f\sqrt{f}},$$

где

$$G_1(x) = \sqrt{2}x; \quad G_2(x) = \frac{2}{3}(x^2-1); \quad G_3(x) = \frac{x^3-7x}{9\sqrt{2}};$$

$$G_4(x) = \frac{6x^4+14x^2-32}{405}; \quad G_5(x) = \frac{9x^5+256x^3-433x}{4860\sqrt{2}}.$$

Некоторые значения $G_i(x)$ приведены в табл. 3.

Аппроксимация 9 (*Гилберта*) [69]. При $f > 30$

$$\chi_p^2(f) = a_0 + a_1 + a_2f^2 + a_3\ln f,$$

где a_i — коэффициенты, приведенные в [69].

Таблица 3

Значения функции $G_i(x)$

P	u_p	$G_1(x)$	$G_2(x)$	$G_3(x)$	$G_4(x)$	$G_5(x)$
0,999	−3,090232	−4,370248	5,699689	−0,619005	−1,602111	−1,273497
0,995	−2,575829	−3,642772	3,756595	0,073889	−0,802517	−0,622767
0,990	−2,326347	−3,289952	2,941261	0,290267	−0,541970	−0,411565
0,975	−1,959964	−2,771807	1,894305	0,486317	−0,272398	−0,194832
0,950	−1,644854	−2,326174	1,137029	0,554981	−0,122957	−0,077898
0,900	−1,281551	−1,812387	0,428249	0,539450	−0,017722	−0,002186
0,750	−0,674490	−0,953873	−0,363376	0,346842	0,060220	0,030881
0,500	0,000000	0,000000	−0,666667	0,000000	0,079012	0,000000
0,250	0,674490	0,953873	−0,353376	−0,346842	0,060220	−0,030881
0,100	1,281551	1,812387	0,428249	−0,539450	−0,017722	0,002186
0,050	1,644854	2,326174	1,137029	−0,554981	−0,122957	0,077898
0,025	1,959964	2,771807	1,894305	−0,486317	−0,272398	0,194832
0,010	2,326347	3,289952	2,941261	−0,290267	−0,541970	0,411565
0,005	2,575829	3,642772	3,756595	−0,073889	−0,802517	0,622767
0,001	3,090232	4,370248	5,699688	0,619005	−1,602111	1,273497

Аппроксимация 10 (*Хоглина*) [70]. При $p > 0{,}05$ ($f \geqslant 30$)

$$\chi_p^2(f) = \left\{1{,}00991\sqrt{f} + 1{,}9518[-\lg(1-p)]^{\frac{1}{2}}\right\}^2;$$

$$\chi_p^2(f) = \left\{1{,}06807\sqrt{f} + 2{,}13161\lg(1-p)^{\frac{1}{2}} - 0{,}04589\sqrt{f}[-\lg(1-p)]^{\frac{1}{2}} - 1{,}37266\right\}^2;$$

$$\chi_p^2(f) = \left\{\sqrt{f} + 2[-\lg(1-p)]^{\frac{1}{2}} - \frac{7}{6}\right\}^2.$$

При $p \leqslant 0{,}05$ ($f \geqslant 30$)

$$\chi_p^2(f) = \left\{1{,}14309\sqrt{f} - 0{,}9459(-\lg p)^{\frac{1}{2}} - 0{,}1318\sqrt{f}(-\lg p)^{\frac{1}{2}} - 0{,}06198\right\}^2;$$

$$\chi_p^2(f) = \left\{0{,}97657\sqrt{f} - 1{,}46049(-\lg p)^{\frac{1}{2}} + 0{,}59025\right\}^2;$$

$$\chi_p^2(f) = \left\{\sqrt{f} - 1{,}5(-\lg p)^{\frac{1}{2}} + 0{,}6\right\}^2.$$

Отличительной чертой этих аппроксимаций является то, что они не требуют знания квантилей стандартного нормального распределения. Квантили χ_p^2 аппроксимируются непосредственно по значениям f и p.

Аппроксимация 11 (*Арояна*) [71]

$$\chi_p^2(f) = f + \sqrt{2}f\left[u_p + b_1\sqrt{\frac{8}{f}} + b_2\frac{8}{f}\right],$$

где b_1 и b_2 — коэффициенты, приведенные в [71].

Аппроксимация 12 (*Пейзера–Пратта*) [72–74]

$$u_p = \frac{d\left[(f-1)\ln\left(\dfrac{f-1}{\chi_p^2(f)}\right) + \chi_p^2(f) - (f-1)\right]^2}{\left|\chi_p^2(f) - (f-1)\right|},$$

где $d = \chi_p^2(f) - f + \dfrac{2}{3} - \dfrac{0{,}08}{f}$; $\chi_p^2(f) \neq f - 1$.

Аппроксимация 13 [36]. Для $p = 0{,}95$ имеет место

$$\chi^2_{0,95}(f) = 112{,}6 - \frac{1}{200}(147{,}4 - f)^2.$$

При $5 \leqslant f \leqslant 30$ погрешность $\leqslant 1\%$.

Аппроксимация 14 [114]. Является уточнением аппроксимации 3 (Вилсона–Хилферти)

$$\chi^2_p(f) = f\left(1 - \frac{2}{9f} + u_p\sqrt{\frac{2}{9f}}\right)^3 \left(\frac{t_p(f)}{u_p}\right)^{\varphi},$$

где $t_p(f)$ — p-квантиль распределения Стьюдента (см. раздел 1.1.9) и

$$\varphi = \begin{cases} 0{,}4(p - 0{,}7) & \text{для } 0{,}7 \leqslant p \leqslant 0{,}825, \\ 0{,}3(0{,}99 - p) & \text{для } p > 0{,}825. \end{cases}$$

Из приведенных аппроксимаций наиболее точны аппроксимации 5, 6 и 8 (уже при $f > 3$ погрешность не превышает 0,05%) [75].

Отметим (ранее мы уже об этом говорили), что через функцию распределения χ^2 могут быть выражены функции многих других распределений: гамма-распределения (см. раздел 1.1.6), распределения Эрланга, экспоненциального распределения (см. раздел 1.1.4), распределения Вейбулла (см. раздел 1.1.5), распределения Пуассона (см. раздел 1.2.2), распределения Рэлея (см. раздел 1.1.15), распределения Паскаля (см. раздел 1.2.4).

Задача 13. *Вычислить различные аппроксимации квантилей* $\chi^2_p(f)$ *при* $p = 0{,}05$ *и* $0{,}95$ *и* $f = 10$ *и оценить ошибку аппроксимации* (*табличные значения* $\chi^2_{0,05}(10) = 3{,}9403$ *и* $\chi^2_{0,95}(10) = 18{,}307$). *Для вычислений понадобятся значения* $u_{0,05} = -1{,}644854$ *и* $u_{0,95} = 1{,}644854$ (*их можно аппроксимировать, пользуясь аппроксимациями для нормального распределения — см. раздел* 1.1.1).

Аппроксимация 1. Имеем

$$\chi^2_{0,05}(10) = 10 + \sqrt{2 \cdot 10} \cdot (-1{,}644854) = 2{,}643989 \quad (\delta = 32{,}8\%);$$
$$\chi^2_{0,95}(10) = 10 + \sqrt{2 \cdot 10} \cdot (1{,}644854) = 17{,}356 \quad (\delta = 5{,}1\%).$$

Большая погрешность объясняется тем, что аппроксимация 1 дает удовлетворительный результат только при $f > 200$.

Аппроксимация 2:

$$\chi^2_{0,05}(10) = \frac{1}{2} \cdot (-1{,}644854 + \sqrt{20 - 1})^2 = 3{,}683 \quad (\delta = 6{,}5\%);$$
$$\chi^2_{0,95}(10) = \frac{1}{2} \cdot (1{,}644854 + \sqrt{20 - 1})^2 = 18{,}022 \quad (\delta = 1{,}6\%).$$

Погрешность аппроксимации остается значительной, так как аппроксимация 2 удовлетворительна только при $f > 100$.

Аппроксимация 3:

$$\chi^2_{0,95}(10) = 10\left[1 - \frac{2}{9 \cdot 10} + 1{,}644854\sqrt{\frac{2}{9 \cdot 10}}\right]^3 = 18{,}29178 \quad (\delta = 0{,}08\%).$$

Аппроксимация 4. Имеем

$$\left[\chi^2_p(f)\right]^{\frac{1}{4}} \approx (f - 0{,}5)^{\frac{1}{4}} + \sqrt{\frac{1}{8\sqrt{f}}}\, u_p.$$

Тогда

$$\chi^2_{0,05}(10) = \left[(10-0{,}5)^{\frac{1}{4}} - 1{,}644854\sqrt{\frac{1}{8\sqrt{10}}}\right]^4 = (1{,}755621543 - 0{,}327026059)^4 = 4{,}16521;$$

$$\chi^2_{0,95}(10) = (1{,}755621543 + 0{,}327026059)^4 = 18{,}81322 \quad (\delta = 2{,}7\%).$$

Аппроксимация 5. Для $p = 0{,}05$ и $x = u_{0,5} = -1{,}644854$ имеем

$$\chi^2_{0,05}(10) = 10\Bigg[1{,}0000886 - \frac{0{,}2237368}{10} - \frac{0{,}01513904}{100} +$$
$$+ \frac{1}{\sqrt{10}}(-1{,}644854)\left(0{,}4713941 + \frac{0{,}02607083}{10} - \frac{0{,}08986007}{100}\right) +$$
$$+ \frac{1}{10}(-1{,}644854)^2\left(0{,}0001348028 + \frac{0{,}01128186}{10} + \frac{0{,}02277679}{100}\right) +$$
$$+ \frac{1}{10\sqrt{10}}(-1{,}644854)^3\left(-0{,}008553069 - \frac{0{,}01153761}{10} - \frac{0{,}01323293}{100}\right) +$$
$$+ \frac{1}{100}(-1{,}644854)^4\left(0{,}00312558 + \frac{0{,}005169654}{10} - \frac{0{,}006950356}{100}\right) +$$
$$+ \frac{1}{100\sqrt{10}}(-1{,}644854)^5\left(-0{,}0008426812 + \frac{0{,}00253001}{10} + \frac{0{,}001060438}{100}\right) +$$
$$+ \frac{1}{1000}(-1{,}644854)^6\left(0{,}00004780499 + \frac{0{,}001450117}{10} + \frac{0{,}001565326}{100}\right)\Bigg]^3 = 3{,}940420741$$
$$(\delta = 0{,}003\%).$$

При $p = 0{,}95$ и $u_{0,95} = 1{,}1644854$ имеем

$$\chi^2_{0,95}(10) = \big(0{,}977563529 + 0{,}52014855 \cdot 0{,}473911322 + 0{,}27554468 \cdot 1{,}4907559 \cdot 10^{-3} -$$
$$-0{,}40728579 \cdot 9{,}8391593 \cdot 10^{-3} + 0{,}07319972 \cdot 3{,}5730418 \cdot 10^{-3} - 0{,}038074725 \cdot 5{,}7907562 \cdot 10^{-4} -$$
$$- 0{,}01980451 \cdot 3{,}155345 \cdot 10^{-5}\big)^3 = 18{,}30737 \quad (\delta = 0{,}002\%).$$

Аппроксимация 6. Для $p = 0{,}05$ имеем

$$h_{10} = \frac{2}{27 \cdot 10}\left(\frac{2\sqrt{2}(1{,}644854^2 - 1)}{3\sqrt{10}} - \frac{(-1{,}644854)^3 - 3(-1{,}644854)}{4}\right) = -2{,}869712417 \cdot 10^{-3}$$

и

$$\chi^2_{0,05}(10) = 10\left[1 - \frac{2}{9 \cdot 10} + (-1{,}644854 + 2{,}8697122417 \cdot 10^{-3})\sqrt{\frac{2}{9 \cdot 10}}\right]^3 = 3{,}93481$$
$$(\delta = 0{,}047\%).$$

Для $\chi^2_{0,95}(10)$ имеем

$$h_{10} = -7{,}4074074 \cdot 10^{-3}(0{,}508495179 + 0{,}121084002) = -4{,}6635494 \cdot 10^{-3}$$

и

$$\chi^2_{0,95}(10) = 10\left[0{,}977777 + (1{,}644854 + 4{,}6635494 \cdot 10^{-3})\sqrt{\frac{2}{9 \cdot 10}}\right]^3 = 18{,}3229961$$
$$(\delta = 0{,}087\%).$$

В модифицированной форме для $p = 0{,}05$

$$h'_{10} = \frac{(9 \cdot 10)(-1{,}644854)^3 - 3(-1{,}644854) - 24(1{,}644854^2 - 1)\sqrt{20}}{486 \cdot 10^2} = -2{,}7102602 \cdot 10^{-3};$$

$$\chi^2_{0,05}(10) = 10\left[1 - \frac{2}{90} + (-1{,}644854 + 2{,}7102602\cdot 10^{-3})\sqrt{\frac{2}{90}}\right]^3 = 3{,}93802 \quad (\delta = 0{,}058\%).$$

Для $p = 0{,}95$ получаем

$$h'_{10} = \frac{-51{,}33961699 - 183{,}0582646}{48600} = -4{,}8230016;$$

$$\chi^2_{0,95}(10) = 10(0{,}977777 + 0{,}24448138)^3 = 18{,}25954 \quad (\delta = 0{,}29\%).$$

Аппроксимация 7:

$$\chi^2_{0,05}(10) = \frac{12\cdot 10\left[\dfrac{5(-1{,}644854)}{6\left(1-\dfrac{1}{18\cdot 10}\right)\sqrt{2\cdot 10}} + 1 - \dfrac{5}{18\cdot 10}\left(1+\dfrac{7}{48\cdot 10}\right)\right]^{\frac{13}{5}} + 10}{13} =$$

$$= \frac{120(-0{,}308212739 + 1 - 0{,}02818287)^{\frac{13}{5}} + 10}{13} = 3{,}94758 \quad (\delta = 0{,}18\%);$$

$$\chi^2_{0,95}(10) = \frac{120(0{,}308212739 + 1 - 0{,}02818287)^{\frac{13}{5}} + 10}{13} = 18{,}30844 \quad (\delta = 0{,}08\%).$$

Аппроксимация 8. Упрощенная формула:

$$\chi^2_{0,05}(10) = 10 + (-1{,}644854)\sqrt{2\cdot 10} + \frac{2}{3}(1{,}644854^2 - 1) +$$
$$+ \frac{1}{9\sqrt{20}}\big(1{,}6448542^2 - 7\cdot(-1{,}644854)\big) = 4{,}134 \quad (\delta = 4{,}9\%);$$

$$\chi^2_{0,95}(10) = 10 + 7{,}356010714 + 1{,}137029788 + 0{,}353286881 = 18{,}846327 \quad (\delta = 2{,}9\%).$$

Точная формула ($p = 0{,}05$ и $u_p = -1{,}644854$):

$$G_1(x) = -2{,}326174; \quad G_2(x) = 1{,}137029; \quad G_3(x) = 0{,}554981;$$
$$G_4(x) = -0{,}122957; \quad G_5(x) = -0{,}002186.$$

$$\chi^2_{0,05}(10) = 10 - 2{,}326174\sqrt{10} + 1{,}137029 + \frac{1}{\sqrt{10}}0{,}554981 - \frac{1}{10\sqrt{10}}0{,}002186 = 3{,}9441565$$
$$(\delta = 0{,}01\%).$$

При $p = 0,95$ и $u_p = 1,644854$ имеем

$$G_1(x) = 2{,}326174; \quad G_2(x) = 1{,}137029; \quad G_3(x) = -0{,}554981;$$
$$G_4(x) = -0{,}122957; \quad G_5(x) = 0{,}077898$$

и

$$\chi^2_{0,95}(10) = 10 + 2{,}326174\sqrt{10} + 1{,}137029 - \frac{1}{\sqrt{10}}0{,}554981 - \frac{1}{10}0{,}122957 +$$
$$+ \frac{1}{10\sqrt{10}}0{,}077898 = 18{,}30770 \quad (\delta = 0{,}004\%).$$

Аппроксимация 10. Для $p = 0{,}05$

$$\chi^2_{0,05}(10) = \left[1{,}14309\sqrt{10} - 0{,}9459(-\lg 0{,}05)^{\frac{1}{2}} - 0{,}13138\sqrt{10}(-\lg 0{,}05)^{\frac{1}{2}} - 0{,}06198\right]^2 =$$
$$= 3{,}9999 \quad (\delta = 1{,}5\%);$$

$$\chi^2_{0,05}(10) = \left[0{,}9765\sqrt{10} - 1{,}46049(-\lg 0{,}05)^{\frac{1}{2}} + 0{,}59025\right]^2 = 4{,}04951 \quad (\delta = 2{,}8\%);$$

$$\chi^2_{0,05}(10) = \left[\sqrt{10} - 1{,}5(-\lg 0{,}05)^{\frac{1}{2}}\right]^2 = 4{,}20798 \quad (\delta = 6{,}8\%).$$

Для $p = 0{,}95$

$$\chi^2_{0,95}(10) = \left\{1{,}0099\sqrt{10} + 1{,}9518\,[-\lg_{10}(1-0{,}95)]^{\frac{1}{2}}\right\}^2 = 29{,}37488 \quad (\delta = 60\%);$$

$$\chi^2_{0,95}(0) = \Big\{1{,}06807\sqrt{10} + 2{,}13161[-\lg_{10}(1-0{,}95)]^{\frac{1}{2}} -$$

$$- 0{,}04589\sqrt{10}[-\lg_{10}(1-0{,}95]^{\frac{1}{2}} - 1{,}97266\Big\}^2 = 21{,}06223;$$

$$\chi^2_{0,95}(10) = \left\{10 + 2\,[-\lg_{10}(1-0{,}95)]^{\frac{1}{2}} - \frac{7}{6}\right\}^2 = 21{,}435 \quad (\delta = 1{,}5\%).$$

Эти аппроксимации явно неточны для малого количества степеней свободы (их рекомендуется применять при $f > 30$).

Аппроксимация 12. Здесь, располагая точными значениями $\chi^2_p(f)$, можно проверить их нормальную аппроксимацию.

Для $p = 0{,}05$ и $f = 10$ имеем

$$d = \chi^2_{0,05}(10) - 10 + \frac{2}{3} - \frac{0{,}08}{10} = -5{,}4010333;$$

$$u_{0,05} = \frac{-5{,}4010333\left(9\ln\dfrac{9}{3{,}9403} + 3{,}9403 - 9\right)^{\frac{1}{2}}}{|3{,}9403 - 9|} = -1{,}644725.$$

Для $p = 0{,}95$ имеем

$$d = 18{,}307 - 10 + \frac{2}{3} - \frac{0{,}08}{10} = 8{,}965666;$$

$$u_{0,95} = \frac{8{,}965666\left(9\ln\dfrac{9}{18{,}307} + 18{,}307 - 9\right)^{\frac{1}{2}}}{18{,}307 - 9} = 1{,}6454169 \quad (\delta = 0{,}02\%).$$

Аппроксимация 13

$$\chi^2_{0,95}(10) = 112{,}6 - \frac{(147{,}4-10)^2}{200} = 18{,}2062 \quad (\delta = 0{,}5\%).$$

Аппроксимация 14. Будем использовать результаты, полученные при применении аппроксимации 3, а именно $\chi^2_{0,05}(10) = 3{,}93152$ и $\chi^2_{0,95}(10) = 18{,}29178$.

Вычислим корректирующий множитель для $p = 0{,}95$:

$$\frac{t_p(f)}{u_p} = \frac{t_{0,95}(10)}{u_{0,95}} = \frac{1{,}812}{1{,}645} = 1{,}10151975,$$

и для $\varphi = 0{,}3(0{,}99 - 0{,}95) = 0{,}012$ имеем $\left(\dfrac{t_{0,95}(10)}{u_{0,95}}\right)^{0,012} = 1{,}00116.$

Тогда

$$\chi^2_{0,95}(10) = 18{,}29178 \cdot 1{,}100116 = 18{,}313,$$

что соответствует ошибке $\delta = 0{,}03\%$, т. е. ошибка по сравнению с аппроксимацией Вилсона–Хилферти снижается более чем в 2 раза.

1.1.9. Распределение Стьюдента (t-распределение)

Описание, применение. Впервые предложено английским статистиком Госсетом (псевдоним Стьюдент) в 1908 г. Если y — нормально распределенная случайная величина с нулевым средним и единичной дисперсией, а независимая от нее случайная величина χ^2 имеет распределение хи-квадрат (см. раздел 1.1.8) с f степенями свободы, то случайная величина $t = y\left(\chi^2/f\right)^{-\frac{1}{2}}$ подчиняется распределению Стьюдента с f степенями свободы. Распределение Стьюдента широко применяется

в задачах обработки экспериментальных данных (например, при построении доверительных интервалов и проверке гипотез относительно среднего при неизвестной дисперсии). С помощью распределения Стьюдента описываются распределения коэффициентов корреляции и регрессии.

Свойства

Обозначение	$t(f)$
Параметр	f — число степеней свободы
Плотность	$\varphi(t;f)=\dfrac{\Gamma\left(\dfrac{f+1}{2}\right)}{\Gamma\left(\dfrac{f}{2}\right)\sqrt{\pi f}}\left(1+\dfrac{t^2}{f}\right)^{-\frac{f+1}{2}}, \quad -\infty<t<\infty$
Функция распределения	$F(t;f)=\mathbf{P}[t(f)<t]=\dfrac{1}{\sqrt{\pi f}}\int\limits_{-\infty}^{t}\left(1+\dfrac{y^2}{f}\right)^{-\frac{f+1}{2}}dy$
Среднее	$\mathbf{M}(t)=0$
Дисперсия	$\mathbf{D}(t)=\dfrac{f}{f-2}, \quad f>2$
Коэффициент вариации	$v=0$
Коэффициент асимметрии	$\alpha_3=0$
Коэффициент эксцесса	$\alpha_4=\dfrac{3(f-2)}{f-4}, \quad f>4$
Медиана	$\mathbf{Me}=0$
Мода	$\mathbf{Mo}=0$

При $f\to\infty$ t-распределение совпадает со стандартным нормальным (хорошая аппроксимация достигается уже при $f>30$). Таблицы распределения Стьюдента можно найти во многих руководствах по математической статистике [7, 23, 24, 25, 56, 57]. График плотности t-распределения напоминает по форме плотность нормального распределения, но значительнее медленнее приближается к оси абсцисс.

Приведем аппроксимации для расчетов, связанных с t-распределением.

Аппроксимация 1

$$t_p(f)=u_p$$

при $f>30$ (u_p — квантиль стандартного нормального распределения).

Аппроксимация 2 (*Корниша–Фишера*) [76]

$$t_p(f)=u_p\left\{1+\frac{u_p^2+1}{4f}+\frac{5u_p^4+16u_p^2+3}{96f^2}+\frac{3u_p^6+19u_p^4+17u_p^2+15}{384f^3}\right\}.$$

При $f\geqslant 5$ погрешность $\leqslant 10^{-3}$.

Аппроксимация 3 (*Кёхлера*) [77, 78]

$$t_p(f)=\left[-0{,}0953-\frac{0{,}631}{f+1}+\frac{0{,}81}{\sqrt{-\ln[4p(1-p)]}}+0{,}076(4p\sqrt{f})^{\frac{1}{f}}\right]^{-1}.$$

При $f\geqslant 8$ погрешность $<1{,}4\%$, при $f\geqslant 50$ погрешность $<0{,}6\%$ для диапазона $0{,}00001<p<0{,}2$.

Аппроксимация 4 (*Нельсона*) [79, 80]. Очень простая аппроксимация для $p=0{,}95$

$$\left|t_{0,95}(f)\sqrt{\frac{f-2}{f}}\right|\approx 2, \quad \text{или (что эквивалентно)} \quad t_{0,975}(f)=-t_{0,025}(f)\approx\sqrt{\frac{4f}{f-2}}.$$

Аппроксимация 5 [36]

$$t_{0,975} \approx 1{,}96 + \frac{2{,}5}{f - 1{,}8}.$$

Аппроксимация 6 [81]

$$\mathbf{P}\{t(f) < t_p(f)\} = \frac{1}{2}\left[\sum_{i=0}^{5} c_i |t^i|\right]^{-8}, \quad \text{где} \quad c_i = \frac{\sum\limits_{j=0}^{4} a_{ij} f^{-j}}{1 + \sum\limits_{j=1}^{2} b_{ij} f^{-j}}, \quad t \leqslant 0,$$

$a_{10} = 0{,}09979441;$	$a_{20} = 0{,}04431742;$	$a_{30} = 0{,}009694901;$	$a_{40} = -0{,}0000918228;$
$a_{11} = -0{,}58121;$	$a_{21} = -0{,}2206018;$	$a_{31} = -0{,}1408854;$	$a_{41} = 0{,}03789901;$
$a_{12} = 1{,}390993;$	$a_{22} = 0{,}03317253;$	$a_{32} = 1{,}88993;$	$a_{42} = -1{,}280346;$
$a_{13} = -1{,}222452;$	$a_{23} = 5{,}679969;$	$a_{33} = -12{,}75532;$	$a_{43} = 9{,}249528;$
$a_{14} = 2{,}151185;$	$a_{34} = -12{,}96519;$	$a_{34} = 25{,}77532;$	$a_{44} = -19{,}08115;$

$a_{50} = 0{,}000579602;$	$b_{11} = -5{,}537409;$	$b_{32} = 14{,}3963;$
$a_{51} = -0{,}02763334;$	$b_{12} = 11{,}42343;$	$b_{41} = -2{,}777816;$
$a_{52} = 0{,}4517029;$	$b_{21} = -5{,}166733;$	$b_{42} = 16{,}461132;$
$a_{53} = -2{,}657967;$	$b_{22} = 13{,}49862;$	$b_{51} = -0{,}5657187;$
$a_{54} = 5{,}127212;$	$b_{31} = -4{,}233736;$	$b_{52} = 21{,}83269.$

При $f \geqslant 5$ погрешность $\leqslant 10^{-4}$.

Аппроксимация 7 (*Пейзера–Пратта*) [72, 73]

$$u_p = \left(f - \frac{2}{3}\right)\left\{\frac{\ln\left(1 + \frac{t_p^2(f)}{f}\right)}{f - \frac{5}{6}}\right\}^{\frac{1}{2}}; \quad u_p = \left(f - \frac{2}{3} + \frac{1}{10f}\right)\left\{\frac{\ln\left(1 + \frac{t_p^2(f)}{f}\right)}{f - \frac{5}{6}}\right\}^{\frac{1}{2}}.$$

Аппроксимация 8 (*Уоллиса*) [74]

$$u_p = \frac{8f+1}{8f+3}\left[f \ln\left(1 + \frac{t_p^2(f)}{f}\right)\right]^{\frac{1}{2}} \quad \text{при} \quad f \geqslant 5.$$

Аппроксимация 9 (*Морана*) [82, 84]

$$t_p(f) = u_p\left(1 - \frac{u_p^2 + 1}{4f}\right)^{-1}.$$

При $f \geqslant 5$ погрешность не более 10^{-3}.

Аппроксимация 10 (*Даусона*) [83]

$$t_p(f) = -0{,}5059 - 1{,}26\lg(2p) - 0{,}1093[\lg(2p)]^2 + \exp\left\{\frac{1{,}994 - 2{,}7497\lg(2p)}{f}\right\}.$$

При $0{,}005 \leqslant p \leqslant 0{,}1$ и $3 \leqslant f \leqslant 8$ погрешность составляет $\pm 5\%$, при $0{,}001 \leqslant p \leqslant 0{,}2$ погрешность равна $\pm 8\%$.

Аппроксимация 11 (*Вонга*) [85, 86]

$$t_p(f) = \left\{f\left[\exp\left(\frac{u_p^2}{0{,}9975f - 0{,}445}\right) - 1\right]\right\}^{\frac{1}{2}}.$$

При $p \approx 0{,}025$ и $8 \leqslant f \leqslant 18$ погрешность $\leqslant 0{,}005$.

Аппроксимация 12 (*Локтева*) [114]

$$t_p(f) = u_p \left\{ 1 - (0.325 - 0.022 \ln f) \frac{1 + \ln\left[p(1-p) + 0{,}004 \dfrac{f-5}{f}\right]}{f} \right\}^2 .$$

При $f \geqslant 3$ максимальная ошибка 0.3 для величины $\dfrac{t_p(f)}{u_p}$.

Задача 14. *Вычислить $t_p(f)$ и оценить ошибку аппроксимации при $p = 0{,}05$, $f = 5$ и $p = 0{,}95$, $f = 10$.*

Точные табличные величины $t_{0,05}(5) = -2{,}0150$ и $t_{0,95}(10) = 1{,}8125$. Вспомним, что $u_{0,95} = 1{,}644854 = -u_{0,05}$.

Аппроксимация 2:

$$t_{0,05}(10) = -1{,}644854 \cdot \left\{ 1 + \frac{1{,}644854^2 + 1}{4 \cdot 5} + \frac{5 \cdot 1{,}644854^4 + 16 \cdot 1{,}644854^2 + 3}{96 \cdot 5^2} + \right.$$
$$\left. + \frac{3 \cdot 1{,}644854^6 + 19 \cdot 1{,}644854^4 + 17 \cdot 1{,}644854^2 + 15}{384 \cdot 5^3} \right\} = -2{,}015308 \quad (\delta = 0{,}015\%);$$

$$t_{0,95}(10) = 1{,}644854 \cdot \left\{ 1 + 0{,}092638617 + 8{,}6342265 \cdot 10^{-3} + 6{,}757481 \cdot 10^{-4} \right\} = 1{,}8125445$$
$$(\delta = 0{,}0024\%).$$

Аппроксимация 3:

$$t_{0,05}(5) = \left\{ -0{,}0953 - \frac{0{,}631}{6} + 0{,}81 \cdot [-\ln(4 \cdot 0{,}05 \cdot 0{,}95)]^{-\frac{1}{2}} + 0{,}076 \cdot (4 \cdot 0{,}5 \cdot \sqrt{5})^{\frac{1}{5}} \right\}^{-1} =$$
$$= 2{,}0293 \quad (\delta = 0{,}7\%);$$

$$t_{0,95}(10) = \left\{ -0{,}0953 - \frac{0{,}831}{6} + 0{,}81 \cdot [-\ln(4 \cdot 0{,}95 \cdot 0{,}5)]^{-\frac{1}{2}} + 0{,}076 \cdot (4 \cdot 0{,}95 \cdot \sqrt{10})^{\frac{1}{10}} \right\}^{-1} =$$
$$= 1{,}761817 \quad (\delta = 2{,}8\%)$$

Аппроксимация 4:

$$t_{0,95}(10) = 2 \cdot \sqrt{\frac{10}{8}} = 2{,}236.$$

Табличное значение $t_{0,975}(10) = 2{,}228$.

Аппроксимация 5:

$$t_{0,975}(10) = 1{,}96 + \frac{2{,}5}{10 - 1{,}8} = 2{,}2649.$$

Аппроксимация 6. Пусть $t_p(5) = 2{,}015$. Требуется отыскать значение p, соответствующее этой квантили. Вычисляем

$$c_1 = \frac{0{,}09879441 - 0{,}58121 \cdot \dfrac{1}{5} + 1{,}390993 \cdot \left(\dfrac{1}{5}\right)^2 - 1{,}222452 \cdot \left(\dfrac{1}{5}\right)^3 + 2{,}151185 \cdot \left(\dfrac{1}{5}\right)^4}{1 - 5{,}537409 \cdot \dfrac{1}{5} + 11{,}42343 \cdot \left(\dfrac{1}{5}\right)^2} = 0{,}0936657.$$

По аналогии находим

$$c_2 = 0,0465169; \quad c_3 = -0,00505661; \quad c_4 = -0,000348176; \quad c_5 = 0,00003584898.$$

Окончательно получаем

$$p = \frac{1}{2}\big(1 + 0{,}0936657 \cdot 2{,}015 + 0{,}045169 \cdot 2{,}015^2 - 0{,}005661 \cdot 2{,}015^3 - 0{,}000348176 \cdot 2{,}015^4 +$$
$$+ 0{,}0000358489 \cdot 2{,}015^5\big) = 0{,}05055.$$

Точное значение $p = 0{,}05$.

Аппроксимация 7:

$$u_{0,05} = \left(5 - \frac{2}{3}\right)\left\{\frac{\ln\left(1 + \dfrac{2{,}015^2}{5}\right)}{5^{-\frac{5}{6}}}\right\}^{\frac{1}{2}} = -1{,}6367697 \quad (\delta = 0{,}5\%);$$

$$u_{0,05} = \left(5 - \frac{2}{3} + \frac{1}{10\cdot 5}\right)\cdot 5\left\{\frac{\ln\left(1 + \dfrac{2{,}015^2}{5}\right)}{5^{-\frac{5}{6}}}\right\}^{\frac{1}{2}} = -1{,}644324 \quad (\delta = 0{,}03\%);$$

$$u_{0,95} = \left(10 - \frac{2}{3}\right)\left\{\frac{\ln\left(1 + \dfrac{1{,}8125^2}{10}\right)}{10^{-\frac{5}{6}}}\right\}^{\frac{1}{2}} = 1{,}64300 \quad (\delta = 0{,}1\%);$$

$$u_{0,95} = \left(10 - \frac{2}{3} + \frac{1}{10\cdot 10}\right)\left\{\frac{\ln\left(1 + \dfrac{1{,}8125^2}{10}\right)}{10^{-\frac{5}{6}}}\right\}^{\frac{1}{2}} = 1{,}64476 \quad (\delta = 0{,}0056\%).$$

Аппроксимация 8:

$$u_{0,05} = \frac{8\cdot 5+1}{8\cdot 5+3}\left[5\ln\left(1 + \frac{2{,}015^2}{5}\right)\right]^{\frac{1}{2}} = -1{,}6438427 \quad (f = 5;\ \delta = 0{,}06\%);$$

$$u_{0,95} = \frac{8\cdot 10+1}{8\cdot 10+3}\left[10\ln\left(1 + \frac{1{,}1825^2}{10}\right)\right]^{\frac{1}{2}} = 1{,}64480229 \quad (f = 10;\ \delta = 0{,}003\%).$$

Аппроксимация 9:

$$t_{0,05}(5) = -1{,}644854\left(1 - \frac{1{,}644854^2+1}{4\cdot 5}\right)^{-1} = -2{,}0189125 \quad (\delta = 0{,}2\%);$$

$$t_{0,95}(10) = -1{,}644854\left(1 - \frac{1{,}644854^2+1}{4\cdot 10}\right)^{-1} = 1{,}812788 \quad (\delta = 0{,}01\%).$$

Аппроксимация 10:

$$t_{0,95}(5) = -0{,}5059 - 1{,}26\lg(2\cdot 0{,}05) - 0{,}1093[\lg(2\cdot 0{,}05)]^2 + \\ + \exp\left\{\frac{1{,}1994 - 2{,}7479\lg(2\cdot 0{,}05)}{5}\right\} = -2{,}8476 \quad (\delta = 41\%).$$

Аппроксимация 11:

$$t_{0,05}(5) = \left\{5\left[\exp\left(\frac{1{,}644854^2}{0{,}9975\cdot 5 - 0{,}445}\right) - 1\right]\right\}^{\frac{1}{2}} = -2{,}01758 \quad (\delta = 0{,}12\%);$$

$$t_{0,95}(10) = \left\{10\left[\exp\left(\frac{1{,}644854^2}{0{,}9975\cdot 10 - 0{,}445}\right) - 1\right]\right\}^{\frac{1}{2}} = 1{,}81189675 \quad (\delta = 0{,}03\%).$$

Аппроксимация 12:

$$t_{0,05}(5) = -1{,}644854\left\{1 - (0{,}325 - 0{,}022\cdot\ln 5)\frac{1 + \ln\left(0{,}95\cdot 0{,}05 + 0{,}004\cdot\dfrac{5-5}{5}\right)}{5}\right\}^{2} = -2{,}058;$$

$$t_{0,95}(10) = 1{,}644854 \left\{ 1 - (0{,}325 - 0{,}022 \cdot \ln 10) \frac{1 + \ln\left(0{,}95 \cdot 0{,}05 + 0{,}004 \cdot \dfrac{10-4}{5}\right)}{10} \right\}^2 = 1{,}8308.$$

Из рассмотренного примера видно, что наиболее просты и достаточно точны аппроксимации 3, 4, 5, 7, 8, 9 и 11.

1.1.10. Распределение Фишера (F-распределение)

Описание, применение. Если две независимые случайные величины χ_1^2 и χ_2^2 распределены по закону хи-квадрат соответственно с f_1 и f_2 степенями свободы, то случайная величина $F = \dfrac{\chi_1^2 f_2}{\chi_2^2 f_1}$ имеет распределение Фишера. F-распределение широко применяется при обработке данных (при сравнении дисперсий, анализе корреляций). С помощью F-распределения можно вычислять некоторые дискретные распределения, например, биномиальное.

Свойства

Обозначение $F(f_1, f_2)$

Параметры $f_1,\ f_2$

Плотность

$$f(x; f_1, f_2) = \frac{\Gamma\left(\dfrac{f_1+f_2}{2}\right)}{\Gamma\left(\dfrac{f_1}{2}\right)\Gamma\left(\dfrac{f_2}{2}\right)} \left(\frac{f_1}{f_2}\right)^{\frac{f_1}{2}} \frac{x^{\frac{f_1}{2}-1}}{\left(1 + \dfrac{f_1}{f_2}x\right)^{\frac{f_1+f_2}{2}}}, \quad x \geqslant 0$$

Функция распределения

$$S(x) = P\{F < x\} = \frac{\Gamma\left(\dfrac{f_1+f_2}{2}\right)}{\Gamma\left(\dfrac{f_1}{2}\right)\Gamma\left(\dfrac{f_2}{2}\right)} f_1^{\frac{f_1}{2}} f_2^{\frac{f_2}{2}} \int_0^x y^{\frac{f_1}{2}-1}(f_2 + f_1)^{\frac{f_1+f_2}{2}} dy, \quad x \geqslant 0$$

Среднее $\quad \mathbf{M}[F(f_1, f_2)] = \dfrac{f_2}{f_2 - 2}, \quad f_2 > 2$

Дисперсия $\quad \mathbf{D}[F(f_1, f_2)] = \dfrac{2f_2^2(f_1 + f_2 - 2)}{f_1(f_2-2)^2(f_2-4)}, \quad f_2 > 4$

Коэффициент вариации $\quad v = \left[\dfrac{2(f_1+f_2-2)}{f_1(f_2-4)}\right]^{\frac{1}{2}}, \quad f_2 > 4$

Коэффициент асимметрии $\quad \alpha_3 = \dfrac{2f_1 + f_2 - 2}{f_2 - 6}\left[\dfrac{8(f_2-4)}{f_1+f_2-2}\right]^{\frac{1}{2}}, \quad f_2 > 6$

Мода $\quad \mathbf{Mo} = \dfrac{f_2(f_1-2)}{f_1(f_2+2)}$

В большинстве руководств по математической статистике, например, [24, 29, 87], приводятся квантили F-распределения. Для F-распределения справедливо соотношение $F_p(f_1, f_2) = \dfrac{1}{F_{1-p}(f_2, f_1)}$, что позволяет ограничиться только таблицами для $p < 0{,}5$ или $p > 0{,}5$. При $f_1 = 1$ и $f_2 = \infty$ или $f_1 = \infty$ и $f_2 = 1$ распределение

совпадает с нормальным, а при $f_1 = 1$, $f_2 = \infty$ или $f_2 = 1$, $f_1 = \infty$ — с распределением квадрата случайной величины, имеющей t-распределение Стьюдента с f_2 (f_1) степенями свободы. При $f_2 \to \infty$ F-распределение совпадает с χ^2-распределением при f_1 степенях свободы.

Величина $\dfrac{f_1 F_p(f_1, f_2)}{f_2 + f_1 F_p(f_1, f_2)}$ подчиняется бета-распределению (см. раздел 1.1.7), поэтому функция F-распределения может быть выражена через функцию бета-распределения

$$\mathbf{P}[F_p(f_1, f_2) < x] = I_{\frac{f_1 x}{f_2 + f_1 x}}\left(\frac{f_1}{2}, \frac{f_2}{2}\right).$$

В общем случае случайную величину $F(f_1, f_2)$, подчиняющуюся распределению Фишера, можно интерпретировать как отношение $\dfrac{f_2\gamma_1}{f_1\gamma_2}$, где γ_1 и γ_2 — независимые случайные величины, подчиняющиеся гамма-распределению (см. раздел 1.1.6) с параметрами $f_1/2$ и $f_2/2$. В связи с широким использованием F-распределения применяются различные его аппроксимации и нормализующие преобразования.

Аппроксимация 1

$$F_p(f_1, f_2) = \left\{\frac{1}{\sqrt{f}}\,t_p(f) + \left[1 + \frac{t_p^2(f)}{f}\right]^{\frac{1}{2}}\right\}^2, \quad f_1 = f_2 = f.$$

Аппроксимация 2 (при f_1, $f_2 > 30$) [26]

$$u_p = \frac{-c_1 + c_2 F_p^{\frac{1}{3}}(f_1, f_2)}{\left[f_1^{-1} + f_2^{-1} F_p^{\frac{2}{3}}(f_1, f_2)\right]^{\frac{1}{2}}} = d, \quad \text{или} \quad d' = d\left(1 + 0{,}08 f_2^{-3} d^4\right) = u_p,$$

где

$$c_{1(2)} = \frac{3}{\sqrt{2}} - \frac{2}{\sqrt{3} f_{1(2)}} = 2{,}1213203 - \frac{0{,}47140452}{f_{1(2)}}.$$

Аппроксимация 3 [32]

$$u_p = \frac{\left[\dfrac{f_1(2f_2 - 1)}{f_2} F_p(f_1, f_2)\right]^{\frac{1}{2}} - (2f_1 - 1)^{\frac{1}{2}}}{\left[1 + \dfrac{f_1}{f_2} F_p(f_1, f_2)\right]^{\frac{1}{2}}}.$$

Аппроксимация 4 (*Полсона*) [88]

$$u_p = \frac{F_p^{\frac{1}{3}}(f_1, f_2)\left(1 - \dfrac{2}{9f_2}\right) - \left(1 - \dfrac{2}{9f_1}\right)}{\left[\dfrac{2}{9f_1} + \dfrac{2}{9f_2} F_p^{\frac{2}{3}}(f_1, f_2)\right]^{\frac{1}{2}}}.$$

Отсюда следует, что

$$F_p(f_1, f_2) = 9f_2 \left\{\frac{\dfrac{f_2}{f_1}\left[(9f_1 - 2)(9f_2 - 2) + 3u_p\left\{2f_1(9f_2 - 2)^2 - 36 f_1 f_2 u_p^2\right\}^{\frac{1}{2}}\right]}{(9f_2 - 2)^2 - 18 f_2 u_p^2}\right\}^3.$$

Аппроксимация 5 (*Пейзера–Пратта*) [72, 73]

$$u_p = d\left\{\frac{1 + qg\left(\dfrac{s}{fp}\right) + pg\left(\dfrac{T}{fq}\right)}{\left(f + \dfrac{1}{6}\right)pq}\right\}^{\frac{1}{2}},$$

где $d = S + \dfrac{1}{6} - \left(f + \dfrac{1}{3}\right)p + 0{,}04\left(\dfrac{q}{f_2} - \dfrac{p}{f_1} + \dfrac{q - 0{,}5}{f_1 + f_2}\right)$; $S = \dfrac{f_2 - 1}{2}$; $T = \dfrac{f_1 - 1}{2}$; $f = \dfrac{f_1 + f_2 - 2}{2}$; $p = \dfrac{f_2}{f_1 F_p(f_1, f_2) + f_2}$; $q = 1 - p$; $g(x) = (1 - x^2 + 2x\ln x)(1 - x) - 2$.

При $f_1, f_2 \geqslant 4$ погрешность $\leqslant 0{,}001$.

Аппроксимация 6 (*Воглера–Нортона*)

$$F_p(f_1, f_2) = \left\{\frac{\left(1 - \dfrac{2}{9f_1}\right)\left(1 - \dfrac{2}{9f_2}\right) + u_p\left[\dfrac{2}{9f_1}\left(1 - \dfrac{2}{9f_2}\right)^2 + \dfrac{2}{9f_2}\left(1 - \dfrac{2}{9f_1}\right)^2 - \dfrac{4u_p^2}{81f_1f_2}\right]^{\frac{1}{2}}}{\left(1 - \dfrac{2}{9f_2}\right) - \dfrac{2}{9f_2}u_p^2}\right\}^3.$$

Аппроксимация эффективна для $f_1, f_2 \geqslant 30$, если p — не на „хвосте“ распределения. Эта аппроксимация следует из нормализующего преобразования Полсона (аппроксимация 4).

Аппроксимация 7 (*Хейнса*) [89]. Основана на аппроксимации Вилсона–Хилферти для χ^2 (см. раздел 1.1.8, аппроксимация 3) и аппроксимации Гастингса для нормального распределения (см. раздел 1.1.1)

$$F_p(f_1, f_2) = \left\{\frac{-2ab - \left[(2ab)^2 - 4(2x_0^2c - a^2)(2x_0^2d - b^2)\right]^{\frac{1}{2}}}{2(2x_0^2c - d^2)}\right\},$$

где

$$a = 1 - \frac{2}{9f_2}; \quad b = 1 - \frac{2}{9f_1}; \quad c = 1 - a; \quad x_0 = -\frac{a_3}{4a_4} + \frac{E - R}{2};$$

$$R = \left\{\left(\frac{a_3}{2a_4}\right)^2 - \frac{a_2}{a_4} + y_0\right\}^{\frac{1}{2}};$$

$$E = 13\left(\frac{a_3}{2a_4}\right)^2 - R^2 - \frac{2a_2}{a_4} - \left\{\frac{1}{4R}\left[\frac{4a_2a_3}{a_4^2} - \frac{8a_1}{a_4} - \left(\frac{a_3}{a_4}\right)^3\right]\right\}^{\frac{1}{2}}; \quad y_0 = U + \frac{a_2}{3a_4};$$

$$a_1 = 0{,}278393; \quad a_2 = 0{,}230389; \quad a_3 = 0{,}000972; \quad a_4 = 0{,}078108;$$

$$U = \operatorname{sgn}(A)|A|^{\frac{1}{3}} + \operatorname{sgn}(B)|B|^{\frac{1}{3}}; \quad A = -\frac{c_2}{2} + \left[\left(\frac{c_2}{2}\right)^2 + \left(\frac{c_1}{3}\right)^3\right]^{\frac{1}{2}};$$

$$B = -\frac{c_2}{2} - \left[\left(\frac{c_2}{2}\right)^2 + \left(\frac{c_1}{3}\right)^3\right]^{\frac{1}{2}}; \quad c_1 = 51{,}21114354\left[\frac{1}{2}(1 - p)\right]^4 - 54{,}06887755;$$

$$c_2 = -100{,}7003283\left[\frac{1}{2}(1 - p)\right]^{\frac{1}{4}} + 86{,}13944869.$$

Аппроксимация слишком сложна и применяется редко.

Аппроксимация 8 (*Картера*) [90] $F_p(f_1, f_2) = \exp(2z)$, где

$$z = \frac{u_p(h+\lambda)^{\frac{1}{2}}}{h} + \left(\frac{1}{f_2-1} + \frac{1}{f_1-1}\right)\left\{\frac{5}{6} + \lambda - \frac{1}{3}\left(\frac{1}{f_2-1} + \frac{1}{f_1-1}\right)\right\};$$

$$h = \frac{2}{\dfrac{1}{f_2-1} - \dfrac{1}{f_1-1}}; \quad \lambda = \frac{1}{6}(u_p^2 - 3).$$

При $f_1, f_2 \geqslant 30$ и $p \geqslant 0{,}0001$ погрешность $\leqslant 10^{-3}$.

Аппроксимация 9 [9, 91, 92]

$$\lg F_p(f_1, f_2) = a(h-b)^{-\frac{1}{2}} - cg, \quad \text{где} \quad h = \frac{2f_1f_2}{f_1+f_2}; \quad g = \frac{f_2-f_1}{f_1f_2}.$$

Коэффициенты a, b и c приведены в таблице:

p	0,50	0,75	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
a	0	0,5859	1,1131	1,4287	1,7023	2,0206	2,2373	2,6841	2,8580
b	–	0,58	0,77	0,95	1,14	1,40	1,61	2,09	2,30
c	0,290	0,355	0,527	0,681	0,846	1,073	1,250	1,672	1,857

Задача 15. *Вычислить, пользуясь аппроксимациями, значение* $F_{0,9}(10, 12)$.

Для справки — точное значение $F_{0,9}(10, 12) = 2{,}1878$.

Аппроксимация 6. Имеем табличное значение $u_{0,9} = 1{,}281551$.

Вычисляем $1 - \frac{2}{9f_1} = 1 - \frac{2}{90} = 0{,}9777$ и $1 - \frac{2}{9f_2} = 1 - \frac{2}{108} = 9{,}98148$. Окончательно находим

$$F_{0,9}(10, 12) =$$

$$= \left\{\frac{0{,}9777 \cdot 0{,}98148 + 1{,}28155\left[0{,}02222 \cdot 0{,}98148^2 + 0{,}0185185 \cdot 0{,}9777^2 - \frac{4}{81 \cdot 10 \cdot 12}1{,}281155^2\right]^{\frac{1}{2}}}{0{,}98148^2 - 0{,}0185185 \cdot 1{,}28155^2}\right\}^3 =$$

$$= \left(\frac{1{,}21140198}{0{,}932891651}\right)^3 = 2{,}189632439 \quad (\delta = 0{,}03\%).$$

Аппроксимация 8. Имеем $U_{0,9} = 1{,}281551$ и

$$h = \frac{2}{\dfrac{1}{12-1} + \dfrac{1}{10-1}} = 9{,}9; \quad \lambda = \frac{1}{6}(1{,}281551^2 - 3) = -0{,}226271172;$$

$$z = \frac{1{,}281551(9{,}9 - 0{,}226271172)^{\frac{1}{2}}}{9{,}9} + \left(\frac{1}{11} + \frac{1}{9}\right)\left[\frac{5}{6} - 0{,}226271172 - \frac{1}{3}\left(\frac{1}{11} + \frac{1}{9}\right)\right] = 0{,}511656907;$$

$$F_{0,9}(10{,}12) = \exp(2 \cdot 0{,}511656907) = 2{,}782399.$$

Погрешность $\delta = 27\%$ велика, так как малы степени свободы f_1 и f_2.

Аппроксимация 9. Имеем $a = 1{,}1131$; $b = 0{,}77$ и $c = 0{,}527$. Далее находим $h = \frac{2 \cdot 12 \cdot 10}{12+10} = 10{,}90909$ и $g = \frac{12-10}{12 \cdot 10} = 0{,}016666$. Тогда

$$\lg F_{0,9}(10{,}12) = 1{,}1131(1{,}90909 - 0{,}77)^{-\frac{1}{2}} - 0{,}527 \cdot 0{,}016666 = 0{,}340787085$$

и

$$F_{0,9}(10, 12) = 2{,}19173 \quad (\delta = 0{,}18\%).$$

Задача 16. *Вычислить нормализующее преобразование для $F_{0,9}(10,12)$ и оценить точность нормализации.*

Для $p = 0{,}9$ имеем $u_{0,9} = 1{,}2811551$.

Аппроксимация 2. Имеем $F_{0,9}(10,12) = 2{,}1878$ и находим

$$c_1 = 2{,}1213203 - \frac{0{,}471404521}{10} = 2{,}074179848, \quad c_2 = 2{,}1213203 - \frac{0{,}47140452}{12} = 2{,}08203659,$$

$$d = u_p = \frac{-2{,}074179848 + 2{,}08203659 \cdot 2{,}1878^{\frac{1}{3}}}{\left(\frac{1}{10} + \frac{1}{12} \cdot 2{,}1878^{\frac{2}{3}}\right)^{\frac{1}{2}}} = 1{,}28212 \quad (\delta = 0{,}01\%);$$

$$d' = 1{,}28212\left(1 + 0{,}8 \cdot \frac{1}{12} \cdot 1{,}282122^4\right) = 1{,}282282 \quad (\delta = 0{,}06\%).$$

Аппроксимация 3:

$$u_{0,9} = \frac{\left[\frac{10 \cdot (2 \cdot 12 - 1)}{12} \cdot 2{,}1878\right]^{\frac{1}{2}} - (2 \cdot 10 - 1)^{\frac{1}{2}}}{\left(1 + \frac{10}{12} \cdot 2{,}1878\right)^{\frac{1}{2}}} = 1{,}259744 \quad (\delta = 0{,}04\%).$$

Аппроксимация 4:

$$u_{0,9} = \frac{2{,}1878^{\frac{1}{3}}\left(1 - \frac{2}{9 \cdot 12}\right) - \left(1 - \frac{2}{9 \cdot 10}\right)}{\left(\frac{2}{9 \cdot 10} + \frac{2}{9 \cdot 12} \cdot 2{,}1878^{\frac{2}{3}}\right)^{\frac{1}{2}}} = 1{,}282122 \quad (\delta = 0{,}04\%).$$

Оценим квантиль $F_{0,9}(10,12)$, пользуясь этой аппроксимацией. Имеем

$$(9f_1 - 2)(9f_2 - 2) = (9 \cdot 10 - 2)(9 \cdot 12 - 2) = 9328;$$

$$\left[2f_1(9f_2 - 2)^2 + 2f_2(9f_1 - 2)^2 - 36f_2f_2u_p^2\right]^{\frac{1}{2}} = 635{,}2015025;$$

$$(9f_2 - 2)^2 - 18f_2u_p^2 - (9 \cdot 12 - 2)^2 - 18 \cdot 12 \cdot 1{,}281551^2 = 10881{,}24744;$$

$$F_{0,9}(10{,}12) = \left[\frac{1{,}2(9328 + 3 \cdot 1{,}281551 \cdot 635{,}2015025)}{10881{,}24744}\right]^3 = 2{,}187013 \quad (\delta = 0{,}03\%).$$

Аппроксимация 5. Вычисляем

$$p = \frac{12}{10 \cdot 2{,}1878 + 12} = 0{,}354212173; \quad q = 1 - p = 0{,}645787826; \quad S = \frac{12 - 1}{2} = 5{,}5;$$

$$T = \frac{10 - 1}{2} = 4{,}5; \quad f = \frac{10 + 12 - 2}{2} = 1; \quad \frac{S}{fp} = 1{,}552741557; \quad \frac{T}{fq} = 0{,}69682329;$$

$$g(1{,}55271667) = -0{,}145733931; \quad g(0{,}69682329) = 0{,}119889378;$$

$$d = 5{,}5 + \frac{1}{6} - \left(10 + \frac{1}{3}\right) \cdot 0{,}354212173 + 0{,}04\left(\frac{0{,}645787826}{12} - \frac{0{,}354212173}{10} + \frac{0{,}645787826}{10 + 12}\right) = 2{,}007475058;$$

$$u_{0,9} = 2{,}007475058\left\{\frac{1 + 0{,}645787826(-0{,}145733931 + 0{,}3542121173 \cdot 0{,}119988937)}{\left(1 + \frac{1}{6}\right) \cdot 0{,}354212173 \cdot 0{,}645787826}\right\}^{\frac{1}{2}} = 1{,}2819445.$$

Относительная погрешность $\delta = 0{,}03\%$.

1.1.11. Усеченное нормальное распределение

Описание, применение. Если из генеральной совокупности, имеющей нормальное распределение $N(\mu, \sigma)$, изъять все элементы, меньшие или бóльшие определенных граничных значений a_1 и a_2, то образуется совокупность, подчиненная усеченному нормальному распределению. Граничные значения называются точками усечения. На практике могут возникнуть случаи двустороннего и одностороннего усечения. Часто это распределение используется при анализе точности производства [93].

Свойства

Двустороннее усечение

Обозначение $N'(\mu, \sigma, a_1, a_2)$

Параметры μ, σ, a_1, a_2

Плотность
$$\varphi'(x) = \frac{\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{a_2-\mu}{\sigma}\right) - \Phi\left(\frac{a_1-\mu}{\sigma}\right)},$$

где $\varphi(x)$ — плотность вероятностей стандартного нормального распределения $N(0,1)$;

$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}\,dt$ — функция Лапласа; μ, σ — параметры исходного нормального распределения; $a_1 < a_2$ — точки усечения

Среднее $\mathbf{M}(x) = \mu - (\lambda_2 - \lambda_1)\sigma$,

где $\lambda_1 = \frac{\varphi(\xi_1)}{\Phi(\xi_2) - \Phi(\xi_1)}$; $\lambda_2 = \frac{\varphi(\xi_2)}{\Phi(\xi_2) - \Phi(\xi_1)}$; $\xi_1 = \frac{a_1-\mu}{\sigma}$; $\xi_2 = \frac{a_2-\mu}{\sigma}$

Дисперсия $\mathbf{D}(x) = (1 + \lambda_1\xi_1 - \lambda_2\xi_2 - \lambda_2 + \lambda_1)^2\sigma^2$

В случае симметричного усечения, т. е. когда $a_1 - \mu = a_2 - \mu = -a_0$ $(a_1 = -a_2)$, имеет место

$$\varphi'(x) = \frac{\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)}{2\Phi\left(\frac{a_0}{\sigma}\right)}, \quad x \in [\mu - a_1 < x < \mu + a_2]$$

Функция распределения
$$F'(x) = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{a_0-\mu}{\sigma}\right)}{2\Phi\left(\frac{a_0}{\sigma}\right)}, \quad a_1 < x < a_2$$

Среднее (мода, медиана) $\mathbf{M}(x) = \mathbf{Me} = \mathbf{Mo} = a_0$

Дисперсия $\mathbf{D}(x) = (1 - 2\xi\lambda)\sigma^2$, где $\xi = \frac{a_0}{\sigma}$; $\lambda = \frac{\varphi\left(\frac{a_0}{\sigma}\right)}{2\Phi\left(\frac{a_0}{\sigma}\right)}$

Одностороннее усечение $(a_2 = \infty)$

Обозначение $N''(\mu, \sigma, a_1)$

Параметры μ, σ, a_1

Среднее $\mathbf{M}(x) = \mu - \gamma\sigma$, где $\gamma = \frac{\varphi(\xi)}{0{,}5 + \Phi(\xi)}$

Дисперсия $\mathbf{D}(x) = \sigma^2(1 - \xi\gamma - \gamma^2)$

Вид кривых распределения разнообразен в зависимости от соотношения параметров a_0, a_1, a_2 и σ, но во многих случаях кривые будут образованы отрезками нормальной кривой с соответственно увеличенными координатами. Кривые симметричны при $a_1 = -a_2$ и несимметричны в иных случаях.

Доля усечения определяется формулами — слева: $0{,}5-\Phi(\xi_1)$ и справа: $0{,}5-\Phi(\xi_2)$.

Если заданы не границы, а доли усечения, то такие выборки называются не полностью определенными (расчеты для них аналогичны).

Задача 17. *Из нормальной выборки с параметрами* $\mu = 100$ *и* $\sigma = 30$ *изъяты члены с* $x < 70$ *и* $x > 160$ ($a_1 = 70$, $a_2 = 160$). *Как изменятся параметры полученного распределения по сравнению с исходным нормальным?*

Вспомним, что функция Лапласа $\Phi(x)$ связана с функцией $F(x)$ соотношением $\dfrac{\Phi(x)}{2} = F(x) - \dfrac{1}{2}$ (см. раздел 1.1.1). Для расчетов $\Phi(x)$ воспользуемся аппроксимацией 5 из раздела 1.1.1, в соответствии с которой

$$\Phi(x) = \frac{1}{2}\left[1 - \exp\left(-\frac{2x^2}{\pi}\right)\right]^{\frac{1}{2}}.$$

Находим

$$\Phi\left(\frac{160-100}{30}\right) = \Phi(2) = \frac{1}{2}\left[1 - \exp\left(-\frac{8}{\pi}\right)\right]^{\frac{1}{2}} = 0{,}48;$$

$$\Phi\left(\frac{70-100}{30}\right) = \Phi(-1) = -\Phi(1) = \frac{1}{2}\left[1 - \exp\left(-\frac{2}{\pi}\right)\right]^{\frac{1}{2}} - 0{,}3431;$$

$$\Phi(2) - \Phi(-1) = \Phi(2) + \Phi(1) = 0{,}48 + 0{,}3431 = 0{,}8231.$$

Вычисляем среднее:

$$\lambda_1 = \frac{\varphi(\xi_1)}{\Phi(\xi_2) - \Phi(\xi_1)};\quad \lambda_2 = \frac{\varphi(\xi_2)}{\Phi(\xi_2) - \Phi(\xi_1)};\quad \xi_1 = \frac{70-100}{30} = -1;\quad \xi_2 = \frac{160-100}{30} = 2.$$

Имеем

$$\Phi(\xi_2) = \Phi(2) = 0{,}48;\quad \Phi(\xi_1) = \Phi(-1) = -0{,}3431;\quad \Phi(\xi_2) - \Phi(\xi_1) = 0{,}823118.$$

Вспомним, что $\varphi(x) = \dfrac{1}{\sqrt{2\pi}}\exp\left\{-\dfrac{x^2}{2}\right\}$; в нашем случае

$$\varphi(-1) = 0{,}398942\exp\left(1 - \frac{1}{2}\right) = 0{,}24197;\quad \varphi(2) = 0{,}398942\exp\left(-\frac{4}{2}\right) = 0{,}05399;$$

$$\lambda_1 = \frac{0{,}24197}{0{,}8231} = 0{,}29396;\quad \lambda_2 = \frac{0{,}05399}{0{,}8231} = 0{,}06559.$$

Итак,

$$\mathbf{M}(x) = \mu - (\lambda_2 - \lambda_1)\sigma = 100 - (0{,}06559 - 0{,}29396)\cdot 30 = 106{,}851,$$

и

$$\mathbf{D}(x) = (1 + \lambda_1\xi_1 - \lambda_2\xi_2 - \lambda_2 + \lambda_1)^2\sigma^2 =$$
$$= (1 + 0{,}29396\cdot(-1) - 0{,}06559\cdot 2 - 0{,}06559 + 0{,}2939)^2\cdot 30^2 = 580{,}652;\quad \sqrt{\mathbf{D}(x)} = 24{,}0967.$$

1.1.12. Распределение модуля случайной величины, распределенной нормально

Описание, применение. Если случайная величина x распределена нормально $N(\mu,\sigma)$, то случайная величина $y = |x|$ будет иметь распределение модуля. Применяется при анализе допусков на изготовление деталей машино- и приборостроения.

Свойства

Обозначение $|N|(\mu, \sigma)$

Параметры μ, σ

Плотность $\psi(y; \mu, \sigma) = \dfrac{1}{\sigma}\left[\varphi\left(\dfrac{y-\mu}{\sigma}\right) + \varphi\left(\dfrac{y+\mu}{\sigma}\right)\right], \quad y > 0,$

где $\varphi(x)$ — плотность стандартного нормального распределения

Функция распределения $F(y; \mu, \sigma) = \Phi\left(\dfrac{y-\mu}{\sigma}\right) + \Phi\left(\dfrac{y+\mu}{\sigma}\right), \quad y > 0,$

где $\Phi(t) = \dfrac{1}{\sqrt{2\pi}}\int\limits_0^t e^{-\frac{x^2}{2}}dx$ — функция Лапласа

Среднее $\mathbf{M}(y) = 2\left[\mu\Phi\left(\dfrac{\mu}{\sigma}\right) + \sigma\varphi\left(\dfrac{\mu}{\sigma}\right)\right]$

(при $\mu = 0$ $\mathbf{M}(y) = \sqrt{\dfrac{2}{\pi}}\,\sigma \approx 0{,}7979\sigma$)

Дисперсия $\mathbf{D}(y) = \sigma^2 + \mu^2 - [\mathbf{M}(y)]^2$

(при $\mu = 0$ $\mathbf{D}(y) = \dfrac{\pi - 2}{2}\,\sigma^2 \approx 0{,}3634\sigma^2$)

Коэффициент асимметрии $\alpha_3 = \dfrac{\sqrt{2}\cdot(4-\pi)}{(\pi-2)\cdot\sqrt{\pi-2}} \approx 0{,}99527$

Коэффициент эксцесса $\alpha_4 = \dfrac{8(\pi-3)}{(\pi-2)^2} \approx 0{,}8691772$

Вид распределения определяется соотношением между параметрами исходного распределения μ и σ. При $\mu \gg \sigma$ распределения N и $|N|$ практически совпадают.

Задача 18. *Случайная величина x распределена нормально с параметрами $\mu = 10$ и $\sigma = 5$. Найти вероятность того, что модуль случайной величины x не превысит 3, т. е. $\mathbf{P}\{|x| < 3\}$. Вычислить параметры распределения случайной величины $y = |x|$.*

Воспользуемся аппроксимацией $5 — \Phi(x) = \dfrac{1}{2}\left[1 - \exp\left(-\dfrac{2x^2}{\pi}\right)\right]^{\frac{1}{2}}$ (раздел 1.1.1) для вычисления $\Phi(x) = F(x) - 0{,}5$.

Имеем

$$\mathbf{P}(|x| < 3) = F(3; 10, 5) = \Phi\left(\frac{3-10}{5}\right) + \Phi\left(\frac{3+10}{5}\right).$$

Далее вычисляем

$$\Phi\left(\frac{3-10}{5}\right) = \Phi(-1{,}4) = -\Phi(1{,}4) = -\frac{1}{2}\left[1 - \exp\left(-\frac{2\cdot 1{,}4^2}{\pi}\right)\right]^{\frac{1}{2}} = -0{,}422154;$$

$$\Phi\left(\frac{3+10}{5}\right) = \Phi(2{,}6) = \frac{1}{2}\left[1 - \exp\left(-\frac{2\cdot 2{,}26^2}{\pi}\right)\right]^{\frac{1}{2}} = 0{,}496608;$$

$$\mathbf{P}(|x < 3|) = -0{,}422154 + 0{,}496608 = 0{,}074454.$$

Находим

$$\mathbf{M}(|x|) = 2\left[10\cdot\Phi\left(\frac{10}{5}\right) + 5\cdot\varphi\left(\frac{10}{5}\right)\right]; \quad \Phi\left(\frac{10}{5}\right) = \Phi(2) = \frac{1}{2}\left[1 - \exp\left(-\frac{8}{\pi}\right)\right]^{\frac{1}{2}} = 0{,}48;$$

$$\varphi(2) = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{4}{2}} = 0{,}053991; \quad \mathbf{M}(|x|) = 2\cdot(10\cdot 0{,}48 + 5\cdot 0{,}053991) = 10{,}13991;$$

$$\mathbf{D}(|x|) = 10^2 + 5^2 - 10{,}13991^2 = 22{,}18223; \quad \sqrt{\mathbf{D}(|x|)} = 4{,}7098.$$

1.1.13. Распределение, порождаемое нормальной плотностью с линейным дрейфом среднего

Описание, применение. Такое распределение возникает при смещении центра группирования мгновенного нормального распределения и является композицией нормального и равномерного распределений, описывая распределение вероятностей за весь период наблюдения. Применяется при изучении износа режущего инструмента, дрейфа параметров электронных приборов.

Свойства

Пусть дрейф среднего во времени описывается формулой

$$\mu(t) = \mu_0 + 2\lambda\sigma_0 t,$$

где μ_0, σ_0 — параметры исходного распределения (при $t = 0$):

$$\lambda = (2\sigma)^{-1}\big[\mu(t)_{\max} - \mu(t)_{\min}\big].$$

Параметры μ, σ, t

Плотность $\varphi(z,\lambda) = \dfrac{\omega}{2\lambda}[\Phi(z\omega+\lambda) - \Phi(z\omega-\lambda)]$,

где $z = \dfrac{x - M(x)}{D(x)}$ — нормированное значение переменной;

$$\omega = \left(1 + \frac{\lambda^2}{3}\right)^{\frac{1}{2}}; \quad \Phi(y) = \frac{1}{\sqrt{2\pi}}\int\limits_0^y e^{-\frac{t^2}{2}}\,dt \text{ — функция Лапласа}$$

Функция распределения

$$F(z,\lambda) = \frac{1}{2}\left\{1 + \frac{1}{\lambda}\big[(z\omega+\lambda)\Phi(z\omega+\lambda) - (z\omega-\lambda)\Phi(z\omega-\lambda) + \varphi(z\omega+\lambda) - \varphi(z\omega-\lambda)\big]\right\},$$

где $\varphi(y)$ — плотность стандартного нормального распределения

Среднее $\mathbf{M}(x) = \mu_0 + \lambda\sigma_0$

Дисперсия $\mathbf{D}(x) = \sigma_0^2\omega^2$

Коэффициент вариации $v = \dfrac{\mu_0 + \lambda\sigma_0}{\sigma_0\omega}$

Коэффициент асимметрии $\alpha_3 = 0$

Коэффициент эксцесса $\alpha_4 = -\dfrac{6}{5}\lambda^4(3+\lambda^2)^{-2}$

При $\sigma_0 \to 0$ распределение стремится к равномерному (см. раздел 1.1.2), при $\lambda \to 0$ — к нормальному (см. раздел 1.1.1). Таблицы плотности $\varphi(z,\lambda)$ и функции $F(z,\lambda)$ для $\lambda = 3$, 6, 10 и 25 приведены в [93].

Задача 19. *Начальное напряжение зажигания газоразрядного прибора распределено нормально с параметрами $\mu_0 = 1000$ В и $\sigma_0 = 150$ В. В течение срока службы среднее значение напряжения зажигания увеличивается до 1500 В. Определить вероятность того, что напряжение зажигания в течение срока службы будет находиться в интервале $900 \div 1100$ В.*

Имеем

$$\mu(t)_{\max} = 1500; \quad \mu(t)_{\min} = 1000; \quad \lambda = \frac{1500 - 1000}{2\cdot 150} = 1{,}666;$$

$$\omega = \left(1 + \frac{\lambda^2}{3}\right)^{\frac{1}{2}} = 1{,}387; \quad \mathbf{M}(x) = \mu_0 + \lambda\sigma_0 = 1000 + 1{,}666\cdot 150 = 1249{,}9;$$

$$\mathbf{D}(x) = \sigma_0^2\omega^2 = 43284{,}802; \quad \sqrt{\mathbf{D}(x)} = 208{,}05.$$

Далее $z_1 = \dfrac{900 - 1249{,}9}{208{,}5} = -1{,}682$; $z_2 = \dfrac{1100 - 1249{,}9}{208{,}5} = -0{,}72$.

Для вычисления значения $\Phi(x)$ воспользуемся аппроксимацией из раздела 1.1.1

$$\Phi(x) = \frac{1}{2}\left\{1 - \frac{\exp\left[\left(\frac{2}{\pi} - 0{,}1253\right)(-x^2)\right]}{1 + 0{,}1253x^2}\right\}^{\frac{1}{2}}.$$

Имеем для $z_1 = -1{,}682$

$$\Phi(-1{,}682 \cdot 1{,}387 + 1{,}666) = \Phi(-0{,}666334) = \frac{1}{2}\left\{1 - \frac{\exp\left[\left(\frac{2}{\pi} - 0{,}1253(-0{,}666334)^2\right)\right]}{1 + 0{,}1253 \cdot 0{,}666334^2}\right\}^{\frac{1}{2}} = \\ = -0{,}247536883;$$

$$\Phi(-1{,}682 \cdot 1{,}387 - 1{,}666) = \Phi(-3{,}998934) = -0{,}4999766;$$
$$\varphi(-0{,}666334) = 0{,}3135136; \quad \varphi(-3{,}998934) = 1{,}344 \cdot 10^{-4}.$$

Для $z_2 = -0{,}72$ имеем

$$\Phi(-0{,}72 \cdot 1{,}387 + 1{,}666) = \Phi(0{,}66796) = 0{,}248057;$$
$$\Phi(-0{,}72 \cdot 1{,}387 - 1{,}666) = \Phi(-2{,}66524) = -0{,}41313076;$$
$$\varphi(0{,}66796) = 0{,}3191724; \quad \varphi(-2{,}66524) = 0{,}0114394.$$

Находим

$$F(z_1,\lambda) = \frac{1}{2}\left\{1 + \frac{1}{1{,}666}(0{,}666334 \cdot 0{,}247536283 + 3{,}998934 \cdot 0{,}49997666 + \ldots)\right\} = 0{,}0453049;$$

$$F(z_2,\lambda) = \frac{1}{2}\left\{1 + \frac{1}{1{,}666}(0{,}66796 \cdot 0{,}248057 - 2{,}66524 \cdot 0{,}41313076 + 0{,}3191724 - \ldots)\right\} = \\ = 0{,}311692225.$$

Окончательно

$$\mathbf{P}(-1{,}682 < z < -0{,}72) = F(-0{,}72) - F(-1{,}682) = 0{,}311692225 - 0{,}0453049 = 0{,}2663873.$$

Следовательно, при таком дрейфе вероятность того, что напряжение зажигания находится в интервале $900 \div 1000$ В, равна $p \approx 0{,}266$.

Для сравнения найдем вероятность попадания в этот интервал напряжения зажигания при отсутствии дрейфа, т. е. когда $\lambda = 0$.

В этом случае $z_1 = \dfrac{900 - 1000}{150} = -0{,}667$; $z_2 = \dfrac{1100 - 1000}{150} = 0{,}667$,

$$\mathbf{P}(-0{,}667 < z < 0{,}667) = 2F(0{,}667) - 1.$$

Находим с помощью аппроксимации 18 из раздела 1.1.1

$$F(0{,}667) = 1 - \frac{1}{2}\exp(-0{,}717 \cdot 0{,}667 - 0{,}416 \cdot 0{,}667^2) = 0{,}7423.$$

Таким образом, $\mathbf{P}(-0{,}667 < z < 0{,}667) = 0{,}4846$. Следовательно, отсутствие дрейфа существенно (почти в 2 раза) увеличивает вероятность нахождения напряжения зажигания в интервале $900 \div 1100$ В.

1.1.14. Распределение, порождаемое нормальной плотностью с линейным дрейфом среднеквадратического отклонения

Описание, применение. Распределения такого типа встречаются при автоматическом изготовлении деталей, когда за время изготовления партии деталей изменяется рассеяние начального нормального распределения (например, при за-

туплении режущего инструмента, при изменении механических свойств заготовок). Распределение погрешностей при измерениях близко к рассматриваемому, если за время проведения измерений имеет место систематическое смещение точности процесса измерения.

Свойства

Пусть дрейф среднеквадратического отклонения описывается формулой

$$\sigma(t) = \sigma_0(1 + 2\lambda t),$$

где σ_0 — параметр исходного распределения ($t = 0$); $\lambda = (2\sigma_0)^{-1}\left[\sigma(t)_{\max} - \sigma(t)_{\min}\right]$.

Плотность $\varphi(z;\lambda) = \dfrac{\omega}{4\sqrt{6\pi\lambda}}\left\{E_i\left(-\dfrac{z^2\omega^2}{6}\right) - E_i\left(-\dfrac{z^2\omega^2}{6(1+2\lambda)}\right)\right\}$,

где $z = \dfrac{x - \mathbf{M}(x)}{\sqrt{\mathbf{D}(x)}}$ — нормированное значение переменной; $E_i(z) = \displaystyle\int\limits_{-\infty}^{z}\frac{e^u}{u}\,du$ — интегральная функция (табулирована); $\omega = (3 + 6\lambda + 4\lambda^2)^{\frac{1}{2}}$

Функция распределения

$$F(z;\lambda) = \frac{1}{2} + \int\limits_0^z \frac{\omega}{4\sqrt{6\pi\lambda}}\left\{E_i\left(-\frac{z^2\omega^2}{6}\right) - E_i\left(-\frac{z^2\omega^2}{6(1+2\lambda)}\right)\right\}dz$$

Среднее $\mathbf{M}(x) = \mu_0$

Дисперсия $\mathbf{D}(x) = \dfrac{\sigma_0^2}{3}\,\omega^2$

Здесь μ_0 и σ_0 — параметры исходного нормального распределения.

Коэффициент вариации $v = \dfrac{1}{\sigma_0\omega}\,\mu_0\sqrt{3}$

Коэффициент асимметрии $\alpha_3 = 0$

Коэффициент эксцесса $\alpha_4 = 12\lambda^2\dfrac{3(1+\lambda)^2 + \frac{1}{5}\lambda^2}{\left[3(1+\lambda^2)^2 + \lambda^2\right]^2}$

Значения $\varphi(z;\lambda)$ и $F(z;\lambda)$ табулированы в [93] для значений параметра $\lambda = 1$, 3, 6 и 9. Значения $F(z;\lambda)$ для различных z и λ приведены в табл. 4.

Задача 20. *Начальная погрешность измерения подчинена нормальному распределению с параметрами $\mu_0 = 100$ ед. и $\sigma_0 = 20$ ед. В процессе проведения измерений σ_0 линейно увеличивается до 60 ед. Найти вероятность того, что при проведении измерений погрешность не превысит 40 ед.*

Имеем

$$\lambda = \frac{60\cdot 20}{2\cdot 20} = 1;\quad \omega = (3 + 6\cdot 1 + 4\cdot 1^2)^{\frac{1}{2}} = 3{,}60555;\quad \mathbf{M}(x) = \mu_0 = 100;$$

$$\sqrt{\mathbf{D}(x)} = \sqrt{\frac{20^2}{3}\cdot 3{,}60555^2} = 41{,}6333;\quad z = \frac{40 - 100}{41{,}6333} = -1{,}44115.$$

Необходимо найти $\mathbf{P}(x < 40) = F(-1{,}44;1) = 1 - F(1{,}44;1)$. По табл. 4 для $\lambda = 1$ и $z = -1{,}44$ находим, что $F(1{,}44;1) \approx 0{,}93$ и, следовательно, искомая вероятность равна $\mathbf{P}(x < 40) = 1 - 0{,}93 = 0{,}07$.

Если бы дрейфа стандартного отклонения не было, то $\mathbf{P}(x < 40) = F\left(\dfrac{40-100}{20}\right) = F(-3) = 1 - F(3)$. Здесь $F(x)$ — функция стандартного нормального распределения.

Так как $F(3) = 0{,}9977$, то $P(x < 40) = 1 - 0{,}9977 = 0{,}0023$.

Таблица 4

Значения функции $F(z;\lambda)$

z	λ				z	λ			
	1	3	6	9		1	3	6	9
0,00	0,5000	0,5000	0,5000	0,5000	2,50	0,9894	0,9866	0,9856	0,9852
0,05	0,5225	0,5281	0,5234	0,5264	2,55	0,9904	0,9877	0,9867	0,9864
0,10	0,5453	0,5560	0,5653	0,5711	2,60	0,9913	0,9887	0,9877	0,9874
0,15	0,5678	0,5831	0,5969	0,6030	2,65	0,9922	0,9896	0,9887	0,9884
0,20	0,5900	0,6094	0,9624	0,6317	2,70	0,9929	0,9905	0,9896	0,9893
0,25	0,6118	0,6346	0,7505	0,6576	2,75	0,9936	0,9913	0,9904	0,9901
0,30	0,6332	0,6585	0,6745	0,6810	2,80	0,9942	0,9920	0,9912	0,9909
0,35	0,6540	0,6810	0,6965	0,7025	2,85	0,9948	0,9927	0,9919	0,9916
0,40	0,6742	0,7022	0,7168	0,7221	2,90	0,9953	0,9933	0,9925	0,9923
0,45	0,6938	0,7220	0,7355	0,7403	2,95	0,9958	0,9939	0,9932	0,9929
0,50	0,7126	0,7405	0,7527	0,7571	3,00	0,9962	0,9944	0,9937	0,9935
0,55	0,7307	0,7578	0,7688	0,7726	3,05	0,9966	0,9949	0,9942	0,9940
0,60	0,7481	0,7739	0,7837	0,7872	3,10	0,9969	0,9953	0,9947	0,9945
0,65	0,7646	0,7889	0,7977	0,8007	3,15	0,9972	0,9958	0,9952	0,9949
0,70	0,7803	0,8029	0,8107	0,8134	3,20	0,9975	0,9961	0,9956	0,9954
0,75	0,7953	0,8160	0,8228	0,8252	3,25	0,9978	0,9965	0,9959	0,9957
0,80	0,8094	0,8282	0,8342	0,8363	3,30	0,9980	0,9965	0,9963	0,9961
0,85	0,8228	0,8394	0,8448	0,8466	3,35	0,9982	0,9971	0,9966	0,9964
0,90	0,8354	0,8503	0,8548	0,8564	3,40	0,9984	0,9974	0,9969	0,9967
0,95	0,8472	0,8603	0,8642	0,8655	3,45	0,9986	0,9976	0,9972	0,9970
1,00	0,8583	0,8697	0,8729	0,8741	3,50	0,9987	0,9978	0,9974	0,9973
1,05	0,8687	0,8784	0,8812	0,8821	3,55	0,9989	0,9980	0,9977	0,9975
1,10	0,8784	0,8866	0,9889	0,8897	3,60	0,9990	0,9982	0,9979	0,9977
1,15	0,8875	0,8943	0,8961	0,8968	3,65	0,9991	0,9984	0,9981	0,9979
1,20	0,8960	0,9015	0,9029	0,9034	3,70	0,9992	0,9985	0,9981	0,9981
1,25	0,9039	0,9082	0,9093	0,9097	3,75	0,9993	0,9987	0,9984	0,9983
1,30	0,9113	0,9145	0,9153	0,9156	3,80	0,9994	0,9988	0,9985	0,9984
1,35	0,9182	0,9204	0,9203	0,9211	3,85	0,9994	0,9989	0,9987	0,9986
1,40	0,9246	0,9259	0,9262	0,9263	3,90	0,9995	0,9990	0,9988	0,9987
1,45	0,9305	0,9311	0,9311	0,9311	3,95	0,9996	0,9991	0,9989	0,9989
1,50	0,9360	0,9359	0,9358	0,9357	4,00	0,9996	0,9992	0,9990	0,9989
1,55	0,9411	0,9405	0,9401	0,9400	4,05	0,9997	0,9993	0,9991	0,9990
1,60	0,9459	0,9447	0,9442	0,9400	4,10	0,9997	0,9994	0,9992	0,9991
1,65	0,9503	0,9486	0,9480	0,9478	4,15	0,9997	0,9994	0,9993	0,9992
1,70	0,9543	0,9528	0,9516	0,9514	4,20	0,9998	0,9995	0,9993	0,9993
1,75	0,9581	0,9558	0,9550	0,9547	4,25	0.9998	0,9996	0,9994	0,9994
1,80	0,9616	0,9590	0,9581	0,9578	4,30	0,9998	0,9996	0,9995	0,9994
1,85	0,9648	0,9620	0,9611	0,9607	4,35	0,9998	0,9996	0,9995	0,9995
1,90	0,9678	0,9648	0,9638	0,9635	4,40	0,9999	0,9996	0,9996	0,9995
1,95	0,9705	0,9674	0,9664	0,9660	4,45		0,9997	0,9996	0,9996
2,00	0,9730	0,9699	0,9668	0,9684	4,50		0,9997	0,9997	0,9996
2,05	0,9754	0,9721	0,9710	0,9707	4,55		0,9998	0,9997	0,9996
2,10	0,9775	0,9743	0,9713	0,9727	4,60		0,9998	0,9997	0,9997
2,15	0,9795	0,9762	0,9751	0,9747	4,65		0,9998	0,9998	0,9997
2,20	0,9813	0,9781	0,9769	0,9765	4,70		0,9998	0,9998	0,9997
2,25	0,9830	0,9798	0,9785	0,9782	4,75		0,9998	0,9998	0,9998
2,30	0,9845	0,9813	0,9802	0,9798	4,80		0,9999	0,9998	0,9998
2,35	0,9859	0,9828	0,9817	0,9813	4,85			0,9998	0,9998
2,40	0,9872	0,9842	0,9831	0,9827	4,90			0,9998	0,9998
2,45	0,9884	0,9854	0,9844	0,9840	4,95			0,9999	0,9998

1.1.15. Распределение Рэлея

Описание, применение. Применяется для описания распределения неотрицательных случайных величин, являющихся векторной суммой двух нормальных случайных величин с равными дисперсиями. Ему подчиняются погрешности геометрической формы (овальность, конусообразность), ошибки взаимного расположения поверхностей (эксцентриситет, разностенность). Распределение Рэлея широко применяется в радиолокации при оценке погрешности обнаружения объектов.

Свойства

Обозначение	$R(a)$
Параметр	a
Плотность	$f(x;a) = \dfrac{x}{a^2}\exp\left(-\dfrac{x^2}{2a^2}\right), \quad x \geqslant 0$
Функция распределения	$F(x;a) = 1 - \exp\left(-\dfrac{x^2}{2a^2}\right), \quad x \geqslant 0$
Среднее	$\mathbf{M}(x) = \sqrt{\dfrac{\pi}{2}}\, a \approx 1{,}253a$
Дисперсия	$\mathbf{D}(x) = \left(2 - \dfrac{\pi}{2}\right)a = 0{,}429a^2$
Коэффициент вариации	$v = \left(\dfrac{4}{4-\pi}\right)^{\frac{1}{2}} = 1{,}913$
Коэффициент асимметрии	$\alpha_3 = \dfrac{2\sqrt{\pi}(\pi-3)}{\sqrt{(4-\pi)^3}} = 0{,}631$
Коэффициент эксцесса	$\alpha_4 = \dfrac{24\pi - 16\pi^2 - 16}{(4-\pi)^2} = 0{,}245$
Мода	$\mathbf{Mo} = a$
Медиана	$\mathbf{Me} = a(2\ln 2)^{\frac{1}{2}} = 1{,}177a$

Для удобства табуляции на практике применяется нормирование случайной величины $z = x/a$. Векторная сумма двух нормальных случайных величин с нулевым средним и одинаковым стандартным отклонением σ имеет распределение Рэлея с параметром σ.

Задача 21. *Случайная величина распределена по закону Рэлея с параметром $a = 4$. Вычислить вероятность того, что случайная величина не превысит значения $x = 3$, и вычислить 95%-ю квантиль распределения.*

Имеем $\mathbf{P}(x < 3) = 1 - \exp\left(-\dfrac{3^2}{2\cdot 4^2}\right) = 0{,}24516$. Для вычисления 95%-й квантили ($x_{0,95}$) используем равенство $0{,}95 = 1 - \exp\left(-\dfrac{x_{0,95}^2}{2\cdot 16}\right)$; $\ln 0{,}05 = -\dfrac{x_{0,95}^2}{2\cdot 16}$, и окончательно имеем $x_{0,95} = (-2\cdot 16\cdot \ln 0{,}05)^{\frac{1}{2}} = 9{,}790987$, т. е. с вероятностью 0,95 случайная величина не превысит значение 9,971.

1.1.16. Распределение Максвелла

Свойства, применение. Векторная сумма трех нормально распределенных случайных величин с нулевыми средними и равными дисперсиями имеет распределение Максвелла. Распределением Максвелла описывается абсолютная величина скорости движения молекул.

Свойства

Обозначение	$M(x, a)$
Параметр	a
Плотность	$f(x; a) = \dfrac{x^2}{a^3}\left(\dfrac{2}{\pi}\right)^{\frac{1}{2}} \exp\left\{-\dfrac{x^2}{2a^2}\right\}, \quad x \geqslant 0$
Функция распределения	$F(x; a) = 2\left[\Phi\left(\dfrac{x}{a}\right) - \dfrac{x}{a}\varphi\left(\dfrac{x}{a}\right)\right], \quad x \geqslant 0,$

где $\varphi(z) = \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{z^2}{2}\right)$; $\Phi(z) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^z e^{-\frac{t^2}{2}}\,dt$

Среднее	$\mathbf{M}(x) = 2a\left(\dfrac{2}{\pi}\right)^{\frac{1}{2}} = 1{,}596a$
Дисперсия	$\mathbf{D}(x) = a^2\left(3 - \dfrac{8}{\pi}\right) = 0{,}454a^2$
Коэффициент вариации	$v = 2\left(\dfrac{2}{3\pi - 8}\right)^{\frac{1}{2}} = 2{,}369$
Коэффициент асимметрии	$\alpha_3 = \dfrac{2\sqrt{2}\cdot(16 - 5\pi)}{(3\pi - 8)(3\pi - 8)^{\frac{1}{2}}} = 0{,}486$
Коэффициент эксцесса	$\alpha_4 = \dfrac{160\pi - 12\pi^2 - 384}{9\pi^2 - 48\pi + 64} = 0{,}108$
Медиана	$\mathbf{Me} = 1{,}538a$
Мода	$\mathbf{Mo} = \sqrt{2}a = 1{,}414a$

Обычно распределение Максвелла используется в нормированной форме ($z = x/a$).

Задача 22. *Вычислить вероятность того, что значение случайной величины, имеющей распределение Максвелла с параметром $a = 10$, превысит $x = 11$.*

Имеем

$$\mathbf{P}(x > 11) = 1 - F(11; 10) = 1 - 2[\Phi(1{,}1) - 1{,}1\cdot\varphi(1{,}1)].$$

Для вычисления $\Phi(x)$ воспользуемся аппроксимацией 6 из раздела 1.1.1, в соответствии с которой

$$\Phi(x) = \frac{1}{2}\left\{1 - \frac{\exp\left[\left(\frac{2}{\pi} - 0{,}1253\right)(-x^2)\right]}{1 + 0{,}1253x^2}\right\}^{\frac{1}{2}}.$$

Тогда

$$\Phi\left(\frac{11}{10}\right) = \Phi(1{,}1) = \frac{1}{2}\left\{1 - \frac{\exp\left[\left(\frac{2}{\pi} - 0{,}1253\right)(-1{,}1^2)\right]}{1 + 0{,}1253\cdot 1{,}1^2}\right\}^{\frac{1}{2}} = 0{,}364784;$$

$$\varphi\left(\frac{11}{10}\right) = \varphi(1{,}1) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1{,}1^2}{2}} = 0{,}217852.$$

Окончательно имеем $\mathbf{P}(x > 11) = 1 - 2\cdot(0{,}364784 - 1{,}1\cdot 0{,}217852) = 0{,}749707$.

1.1.17. Распределение экстремального значения

Описание, применение. Рассмотрим распределение вероятностей экстремальных выборочных значений случайной величины x: $y_1 = \min(x_1, x_2, \ldots, x_n)$ и $y_n = \max(x_1, x_2, \ldots, x_n)$. Различают распределения экстремальных значений последовательности независимых случайных величин двух типов, когда распределение самих случайных величин сосредоточено на конечном и бесконечном интервалах. Если распределение исходных случайных величин ограничено некоторым интервалом $[a, b]$, то для экстремальных значений имеет место так называемое распределение типа III с функцией

— для минимального значения

$$F_{\mathrm{III},1} = 1 - \exp\left[-\left(\frac{x-a}{\alpha}\right)^{\beta}\right], \quad x > a;$$

— для максимального значения

$$F_{\mathrm{III},n} = \exp\left[-\left(\frac{b-x}{\alpha}\right)^{\beta}\right], \quad x < b.$$

Очевидно, что распределения этого типа совпадают с распределением Вейбулла (см. раздел 1.1.5).

Наибольшее практическое применение находит закон распределения экстремальных значений для выборки случайных величин, распределенных на бесконечном интервале. Этот закон принято называть двойным показательным, или распределением типа I [12, 94]. Рассмотрим распределение минимального значения (распределение максимального значения следует из него заменой знака перед x).

Свойства

Плотность	$f(x; a, b) = \frac{1}{b}\exp\left(\frac{x-a}{b}\right)\exp\left\{-\exp\left(\frac{x-a}{b}\right)\right\}$
Функция распределения	$F(x; a, b) = 1 - \exp\left\{-\exp\left(\frac{x-a}{b}\right)\right\}$
Среднее	$\mathbf{M}(x) = a - 0{,}577b$
Дисперсия	$\mathbf{D}(x) = \frac{\pi^2}{6}, b^2 = 1{,}645b^2$
Коэффициент вариации	$v = \frac{a - 0{,}577b}{1{,}645b^2} = 0{,}608\frac{a}{b^2} - \frac{0{,}351}{b}$
Коэффициент асимметрии	$\alpha_3 = -1{,}14$
Коэффициент эксцесса	$\alpha_4 = 2{,}4$
Мода	$\mathbf{Mo} = a$
Медиана	$\mathbf{Me} = a + b\ln\ln 2 = a - 0{,}366b$

Распределение неограничено и несимметрично. Если случайная величина x имеет распределение Вейбулла, то случайная величина $y = \ln x$ имеет распределение минимального значения. Поэтому необходимые расчеты и оценки, связанные с распределением экстремальных значений, могут быть получены из расчетов и оценок распределения Вейбулла.

Если имеется последовательность n независимых случайных величин, имеющих распределение минимального значения с параметрами a и b, то наименьшая из них имеет такое же распределение с параметрами $a - b\ln n$ и b.

Задача 23. *Известно, что распределение минимальной наработки электронного прибора имеет среднее значение, равное* 1000 *ч, и стандартное отклонение, равное* 750 *ч. Найти вероятность того, что значение наработки превысит* 1500 *ч, и вычислить* 95%*-ю квантиль распределения минимальной наработки.*

Имеем $a = 0{,}577b = 1000$ и $\sqrt{1{,}645}b = 750$.

Отсюда $a = 1000 + 0{,}577 \cdot 584{,}76 = 1337{,}41$ и $b = \dfrac{750}{\sqrt{1{,}645}} = 584{,}76$.

Далее $\mathbf{P}(x > 1500) = 1 - \mathbf{P}(x < 1500) = \exp\left\{-\exp\left(\dfrac{1500 - 1337{,}41}{584{,}76}\right)\right\} = 0{,}266989.$

p-квантиль находим из условия $p = 1 - \exp\left\{-\exp\left(\dfrac{x_p - a}{b}\right)\right\}$, где

$$x_p = a + b\ln[-\ln(1-p)] = 1337{,}41 + 584{,}76\ln(-\ln 0{,}05) = 1979,$$

т. е. вероятность того, что наработка превысит 1979 ч, равна 0,05.

1.1.18. Треугольное распределение (распределение Симпсона)

Описание, применение. Сумма двух независимых равномерно распределенных случайных величин имеет треугольное распределение.

Свойства

Параметры	a, b
Плотность	$f(x;a,b) = \begin{cases} 0, & x \leqslant a; \\ \dfrac{4(x-a)}{(b-a)^2}, & a < x < \dfrac{a+b}{2}; \\ 0, & x \geqslant a. \end{cases}$
Функция распределения	$F(x;a,b) = \begin{cases} 0, & x \leqslant a; \\ \dfrac{2(x-a)^2}{(b-a)^2}, & a < x \leqslant \dfrac{a+b}{2}; \\ 1 - \dfrac{2(b-x)^2}{(b-a)^2}, & \dfrac{a+b}{2} < x < b; \\ 1, & x \geqslant b. \end{cases}$
Среднее	$\mathbf{M}(x) = \dfrac{a+b}{2}$
Дисперсия	$\mathbf{D}(x) = \dfrac{(b-a)^2}{24}$
Коэффициент вариации	$v = \dfrac{1}{\sqrt{b}}\dfrac{b-a}{b+a}$
Коэффициент асимметрии	$\alpha_3 = 0$
Коэффициент эксцесса	$\alpha_4 = 2{,}4$
Мода	$\mathbf{Mo} = \dfrac{a+b}{2}$
Медиана	$\mathbf{Me} = \dfrac{a+b}{2}$

Распределение применяется редко, чаще всего в демонстрационных целях.

Задача 24. *Случайная величина x имеет треугольное распределение с дисперсией $\mathbf{D}(x) = 24$ и со средним $\mathbf{M}(x) = 8$. Найти вероятность того, что значение случайной величины будет находиться в интервале $[0, 5]$.*

Имеем $\mathbf{M}(x) = \dfrac{a+b}{2} = 8$ и $\mathbf{D}(x) = \dfrac{(b-a)^2}{24} = 24$, отсюда $b - a = 24$ и $b = 20$, $a = -4$.

Далее $\mathbf{P}(0 < x < 5) = F(5) - F(0) = \dfrac{2\cdot(5+4)^2}{(20+4)^2} - \dfrac{2\cdot(0+4)^2}{(20+4)^2} = 0{,}22569.$

1.1.19. Распределение Коши

Описание, применение. Описывает распределение отношения двух независимых нормально распределенных случайных величин. В прикладной статистике используется редко.

Свойства

Параметры	a, b
Плотность	$f(x; a, b) = \dfrac{1}{\pi b}\left[\exp\left(\dfrac{x-a}{b}\right)^2 + 1\right]$
Функция распределения	$F(x; a, b) = \dfrac{1}{2} + \dfrac{1}{\pi}\,\mathrm{arctg}\left(\dfrac{x-a}{b}\right)$
Мода	$\mathbf{Mo} = a$
Медиана	$\mathbf{Me} = b$

У распределения Коши не существует ни среднего, ни дисперсии. Отношение двух независимых случайных величин, имеющих стандартное нормальное распределение, имеет распределение Коши с параметрами $a = 0$ и $b = 1$. Распределение Коши совпадает с t-распределением Стьюдента (см. раздел 1.1.9) при $f = 1$ степени свободы.

Сумма n независимых случайных величин, имеющих распределение Коши с параметрами a_i и b_i, также имеет распределение Коши с параметрами $\sum a_i$ и $\sum b_i$. Если случайная величина x имеет распределение Коши с параметрами a и b, то обратная ей случайная величина $y = 1/x$ будет иметь распределение Коши с параметрами $a' = a/(a^2 + b^2)$ и $b' = b/(a^2 + b^2)$.

Задача 25. *Отношение энергии разряда W газоразрядного прибора к его предельной энергии $W_{пр}$ называется фактором нагрузки k. Известно, что предельная энергия прибора распределена нормально со средним 1000 Дж и стандартным отклонением 150 Дж. Также нормально со средним 500 Дж и стандартным отклонением 60 Дж распределена энергия разряда прибора во время работы. Необходимо вычислить вероятность того, что при энергии разряда 600 Дж фактор нагрузки не превысит 0,7. Вычислить вероятность того, что значение k будет лежать в интервале $0{,}4 \div 0{,}5$.*

Запишем отношение двух нормированных нормально распределенных случайных величин

$$y = \frac{\dfrac{600-500}{60}}{\dfrac{W_{\text{пр}} - 1000}{150}} = \frac{\dfrac{600-500}{60}}{\dfrac{\dfrac{600}{k} - 1000}{150}} = \frac{5}{4}\cdot\frac{k}{3-5k}.$$

Очевидно, что требование $k \leqslant 0{,}7$ и $0{,}4 \leqslant k \leqslant 0{,}5$ эквивалентно условиям $y \geqslant -7/4$ и $0{,}5 \leqslant y \leqslant 1{,}25$. Случайная величина y имеет распределение Коши с параметрами $a = 0$, $b = 1$, и

$$\mathbf{P}\left(y \geqslant -\frac{7}{4}\right) = 1 - \mathbf{P}\left(y \leqslant -\frac{7}{4}\right) = 1 - F(x; 0, 1) = 1 - \frac{1}{2} - \frac{1}{\pi}\,\mathrm{arctg}\left(-\frac{7}{4}\right) = 0{,}83475.$$

Далее

$$\mathbf{P}(0{,}5 \leqslant y \leqslant 1{,}25) = F(1{,}25) - F(0{,}5) = \frac{1}{2} + \frac{1}{\pi}\,\mathrm{arctg}\,1{,}25 - \frac{1}{2} - \frac{1}{\pi}\,\mathrm{arctg}\,0{,}5 = 0{,}137639.$$

1.1.20. Логистическое распределение

Описание, применение. Чаще всего используется в описательной статистике. Может быть использовано как простейшая модель для приближения распределения нормальной случайной величины.

Свойства

Параметры	a, b
Плотность	$f(x;a,b) = \dfrac{\exp\left[\dfrac{\pi(x-a)}{\sqrt{3}b}\right]}{\dfrac{\sqrt{3}}{\pi}b\left\{1+\exp\dfrac{\pi(x-a)}{\sqrt{3}b}\right\}^2}$
Функция распределения	$F(x;a,b) = \left\{1+\exp\left[\dfrac{\pi(x-a)}{\sqrt{3}\,b}\right]\right\}^{-1}$
Среднее	$\mathbf{M}(x) = a$
Дисперсия	$\mathbf{D}(x) = b^2$
Коэффициент вариации	$v = \dfrac{a}{b}$
Коэффициент асимметрии	$\alpha_3 = 0$
Коэффициент эксцесса	$\alpha_4 = 4{,}2$
Мода	$\mathbf{Mo} = a$
Медиана	$\mathbf{Me} = a$

Задача 26. *Случайная величина имеет логистическое распределение со средним $a = 15$ и коэффициентом вариации $v = 3{,}3$. Вычислить вероятность того, что значение случайной величины не превысит 10. Найти 15%-ю квантиль распределения.*

Имеем $b = \dfrac{a}{v} = \dfrac{15}{3{,}5} = 4{,}2857$ и

$$P(x < 10) = F(10; 15; 4{,}2857) = \left\{1+\exp\left[-\frac{\pi(10-15)}{\sqrt{3}\cdot 4{,}2857}\right]\right\}^{-1} = 0{,}1075.$$

Далее из условия

$$0{,}15 = \frac{1}{1+\exp\left[-\dfrac{\pi(u_{0,15}-15)}{\sqrt{3}\cdot 4{,}2857}\right]} \quad \text{имеем} \quad u_{0,15} = 15 - \frac{\sqrt{3}\cdot 4{,}2857}{\pi}\ln\frac{0{,}85}{0{,}15} = 10{,}9.$$

1.1.21. Распределение Парето

Описание, применение. Применяется в основном в описательной статистике. Впервые рассмотрено при изучении распределения доходов населения. Иногда используется как простейшая математическая модель изменения интенсивности отказов приборов на этапе приработки.

Свойства

Параметр	c
Плотность	$f(x;c) = cx^{-(c+1)}, \quad 1 \leqslant x \leqslant \infty$
Функция распределения	$F(x;c) = 1 - x^{-c}, \quad 1 \leqslant x \leqslant \infty$
Среднее	$\mathbf{M}(x) = \dfrac{c}{c-1}, \quad c > 1$

Дисперсия $\mathbf{D}(x) = \frac{c}{c-2} - \left(\frac{c}{c-1}\right)^2, \quad c > 2$

Коэффициент вариации $v = \left(\frac{c-1}{c}\right)^{\frac{1}{2}}, \quad c > 2$

Если случайная величина x распределена равномерно на интервале $[0, 1]$, то случайная величина $(1/x)^{\frac{1}{c}}$ имеет распределение Парето с параметром c.

Задача 27. *Случайная величина x имеет распределение Парето с параметром $c = 2{,}4$. Вычислить вероятность того, что $x \leqslant 5$, и 95%-ю квантиль распределения.*

Имеем $\mathbf{P}(x \leqslant 5) = F(5; 2{,}4) = 1 - 5^{-2{,}4} = 0{,}979$. Из условия $0{,}95 = 1 - u_{0,95}^{-2,4}$ получаем $u_{0,95} = 0{,}05^{-\frac{1}{2,4}} = 3{,}484$.

1.1.22. Композиции законов распределения вероятностей случайных величин, возникающие при расчете надежности по схеме „нагрузка–напряжение“

Если x — прочность объекта испытаний, а y — действующая на него нагрузка, то вероятность безотказной работы объекта испытаний равна

$$R = \mathbf{P}(x > y) = \mathbf{P}(x - y > 0).$$

В общем случае x и y являются случайными величинами с плотностями распределения вероятностей $f(x)$ и $f(y)$ соответственно. Легко видеть, что вероятность безотказной работы равна $\mathbf{P}(x - y > 0) = 1 - F(x - y)$, где $F(x - y)$ — функция распределения разности случайных величин x и y.

Рассмотрим соотношения для вероятности безотказной работы R при различных законах распределения вероятностей значений прочности и напряжения.

Случай 1. Прочность и напряжение распределены нормально (см. раздел 1.1.1)

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]; \quad f(y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_y}{\sigma_y}\right)^2\right].$$

Тогда

$$R = \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{\mu_x-\mu_y}{\sqrt{\sigma_x^2+\sigma_y^2}}}^{\infty} e^{-\frac{z^2}{2}} dz = 1 - F\left(-\frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}\right),$$

где $F(z)$ — функция стандартного нормального распределения.

Случай 2. Прочность и напряжение распределены логарифмически нормально (см. раздел 1.1.3):

$$f(x) = \frac{1}{x\sigma_x\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_x^2}(\ln x - \mu_x)^2\right\};$$

$$f(y) = \frac{1}{y\sigma_y\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_y^2}(\ln y - \mu_y)^2\right\}; \quad R = 1 - F\left(-\frac{\mu_{\ln x} - \mu_{\ln y}}{\sqrt{\sigma_{\ln x}^2 + \sigma_{\ln y}^2}}\right),$$

где $\mu_{\ln x}$, $\mu_{\ln y}$, $\sigma_{\ln x}$, $\sigma_{\ln y}$ — параметры нормальных распределений величин $\ln x$ и $\ln y$.

Случай 3. Прочность и напряжение распределены по экспоненциальному закону (см. раздел 1.1.4)

$$f(x) = \lambda_x \exp(-\lambda_x x); \quad f(y) = \lambda_y \exp(-\lambda_y y); \quad R = \frac{\lambda_y}{\lambda_x + \lambda_y}.$$

Случай 4. Прочность распределена нормально (экспоненциально), а напряжение экспоненциально (нормально):

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}; \quad f(y) = \lambda_y \exp\{-\lambda_y y\}.$$

Тогда [93]

$$R = 1 - F\left(-\frac{\mu_x}{\sigma_x}\right) - \exp\left\{-\frac{1}{2}\left(2\mu_x\lambda_y - \lambda_y^2\sigma_x^2\right)\right\}\left\{1 - F\left(\frac{-\mu_x - \lambda_y^2\sigma_x^2}{\sigma_x}\right)\right\}.$$

При

$$f(x) = \lambda_x \exp\{-\lambda_x x\}; \quad f(y) = \frac{1}{\sigma_y\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}$$

имеем

$$R = F\left(-\frac{\mu_y}{\sigma_y}\right) + \exp\left\{-\frac{1}{2}\left(2\mu_y\lambda_x - \lambda_x^2\sigma_y^2\right)\right\}\left\{1 - F\left(\frac{-\mu_y - \lambda_x^2\sigma_y^2}{\sigma_y}\right)\right\}.$$

Случай 5. Прочность и напряжение имеют гамма-распределение (см. раздел 1.1.6)

$$f(x) = \frac{1}{\alpha_x!\,\beta_x^{\alpha_x+1}}\,x^{\alpha_x}\exp\left(-\frac{x}{\beta_x}\right); \quad f(y) = \frac{1}{\alpha_y!\,\beta_y^{\alpha_y+1}}\,y^{\alpha_y}\exp\left(-\frac{y}{\beta_y}\right).$$

Вероятность безотказной работы имеет вид:
при $\alpha_x \neq 0$ и $\alpha_y \neq 0$

$$R = \frac{\Gamma(\alpha_x + \alpha_y - 2)}{\Gamma(\alpha_x - 1)\Gamma(\alpha_y - 1)}\,\mathrm{B}_{\frac{r}{1+r}}(\alpha_x - 1, \alpha_y - 1),$$

где $\Gamma(\ldots)$ — гамма-функция, $\mathrm{B}_\gamma(a, b)$ — неполная бета-функция и $r = \dfrac{\beta_y}{\beta_x}$;
при $\alpha_x = \alpha_y = 0$ (x и y имеют экспоненциальное распределение)

$$R = \frac{\beta_y}{\beta_y + \beta_x};$$

при $\alpha_x = 0$, $\alpha_y \neq 0$ (прочность распределена экспоненциально, а напряжение имеет гамма-распределение)

$$R = \left(\frac{\beta_y}{\beta_y + \beta_x}\right)^{\alpha_y - 1};$$

при $\alpha_x \neq 0$, $\alpha_y = 0$ (прочность имеет гамма-распределение, а напряжение распределено экспоненциально)

$$R = 1 - \left(\frac{\beta_x}{\beta_x + \beta_y}\right)^{\alpha_x - 1}.$$

Случай 6. Напряжение распределено нормально, а прочность по закону Вейбулла (см. раздел 1.1.5):

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{(x-\mu_x)^2}{\sigma_x^2}\right\}; \quad f(y) = \frac{\beta_y}{\alpha_y^{\beta_y}}\,y^{\beta_y - 1}\exp\left\{-\left(\frac{y}{\alpha_y}\right)^{\beta_y}\right\}.$$

Вероятность отказа равна

$$1 - R = 1 - F\left(-\frac{\mu_x}{\sigma_x}\right) - \frac{1}{\sqrt{2\pi}}\frac{\alpha_y}{\sigma_x}\int\limits_0^\infty \exp\left\{-z^{\beta_y} - \frac{1}{2}\left[\frac{\alpha_y}{\sigma_x}z - \frac{\mu_x}{\sigma_x}\right]^2\right\}dz.$$

Таблицы значений интеграла приведены в [95] и в табл. 5.

Таблица 5

Значения (1 − *R*), умноженные на 10^4 ($A = -\mu_x/\sigma_x$, $C = \alpha_y/\sigma_x$)

A	C									
	10	20	30	40	50	60	70	80	90	100
$\beta_y = 1$										
0,8	0001	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,6	0003	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,4	0004	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,2	0005	0001	0000	0000	0000	0000	0000	0000	0000	0000
0,0	0008	0001	0000	0000	0000	0000	0000	0000	0000	0000
−0,2	0011	0001	0000	0000	0000	0000	0000	0000	0000	0000
−0,4	0016	0002	0001	0000	0000	0000	0000	0000	0000	0000
−0,6	0022	0003	0001	0000	0000	0000	0000	0000	0000	0000
−0,8	0030	0005	0002	0001	0000	0000	0000	0000	0000	0000
−1,0	0041	0005	0002	0001	0000	0000	0000	0000	0000	0000
−1,4	0069	0009	0003	0001	0001	0000	0000	0000	0000	0000
−1,8	0111	0014	0004	0002	0001	0001	0000	0000	0000	0000
−2,2	0169	0022	0006	0003	0001	0001	0001	0000	0000	0000
−2,6	0247	0032	0009	0004	0002	0001	0001	0000	0000	0000
−3,0	0349	0045	0013	0006	0003	0002	0001	0001	0000	0000
−3,4	0475	0062	0018	0008	0004	0002	0001	0001	0001	0000
−3,8	0630	0082	0024	0010	0005	0003	0002	0001	0001	0001
−4,2	0815	0108	0032	0014	0007	0004	0003	0002	0001	0001
−4,6	1031	0138	0041	0017	0009	0005	0003	0002	0002	0001
−5,0	1279	0173	0052	0022	0011	0006	0004	0003	0002	0001
−5,5	1634	0255	0067	0029	0015	0008	0005	0004	0003	0002
−6,0	2037	0287	0086	0036	0019	0011	0007	0005	0003	0002
−6,5	2485	0360	0108	0046	0023	0014	0009	0006	0004	0003
−7,0	2973	0443	0134	0057	0029	0017	0011	0007	0005	0004
−8,0	4039	0645	0196	0083	0043	0025	0016	0010	0007	0005
−9,0	5265	0897	0276	0117	0050	0035	0022	0015	0010	0008
−10,0	6269	1202	0374	0160	0082	0048	0030	0020	0014	0010
$\beta_y = 2$										
0,8	0011	0003	0001	0001	0000	0000	0000	0000	0000	0000
0,6	0017	0004	0002	0001	0001	0000	0000	0000	0000	0000
0,4	0025	0007	0003	0002	0001	0001	0001	0000	0000	0000
0,2	0035	0009	0004	0002	0001	0001	0001	0001	0000	0000
0,0	0049	0012	0006	0003	0002	0001	0001	0001	0001	0001
−0,2	0067	0017	0008	0004	0003	0002	0001	0001	0001	0001
−0,4	0089	0023	0010	0006	0004	0003	0002	0001	0001	0001
−0,6	0116	0030	0013	0007	0005	0003	0002	0002	0001	0001
−0,8	0149	0038	0017	0010	0006	0004	0003	0002	0002	0002
−1,0	0188	0048	0021	0012	0008	0005	0004	0003	0002	0002
−1,4	0284	0073	0032	0018	0012	0008	0006	0005	0004	0003
−1,8	0407	0105	0047	0026	0017	0012	0009	0007	0005	0004
−2,2	0557	0144	0065	0036	0023	0016	0012	0009	0007	0006
−2,6	0733	0191	0086	0048	0031	0022	0016	0012	0010	0008
−3,0	0935	0246	0110	0062	0040	0028	0020	0016	0012	0010
−3,4	1159	0308	0138	0078	0050	0035	0026	0020	0015	0013
−3,8	1406	0377	0170	0096	0062	0043	0031	0024	0019	0015
−4,2	1671	0453	0205	0116	0074	0052	0038	0029	0023	0019
−4,6	1954	0536	0243	0137	0088	0061	0045	0035	0027	0022

Продолжение таблицы 5

A	C									
	10	20	30	40	50	60	70	80	90	100
	$\beta_y = 2$									
−5,0	2251	0626	0284	0161	0103	0072	0053	0041	0032	0026
−5,5	2640	0748	0341	0193	0124	0086	0064	0049	0038	0031
−6,0	3043	0879	0402	0228	0147	0102	0076	0058	0046	0037
−6,5	3457	1020	0468	0266	0171	0119	0088	0067	0053	0043
−7,0	3876	1170	0539	0307	0198	0138	0101	0078	0062	0050
−8,0	4713	1493	0695	0398	0256	0179	0132	0101	0080	0065
−9,0	5525	1845	0969	0499	0322	0225	0166	0127	0101	0082
−10,0	6285	2222	1059	0611	0396	0277	0204	0157	0124	0100
	$\beta_y = 3$									
0,8	0001	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,6	0003	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,4	0004	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,2	0005	0001	0000	0000	0000	0000	0000	0000	0000	0000
0,0	0008	0001	0000	0000	0000	0000	0000	0000	0000	0000
−0,2	0011	0001	0000	0000	0000	0000	0000	0000	0000	0000
−0,4	0016	0002	0001	0000	0000	0000	0000	0000	0000	0000
−0,6	0022	0003	0001	0000	0000	0000	0000	0000	0000	0000
−0,8	0030	0005	0002	0001	0000	0000	0000	0000	0000	0000
−1,0	0041	0005	0002	0001	0000	0000	0000	0000	0000	0000
−1,4	0069	0009	0003	0001	0001	0000	0000	0000	0000	0000
−1,8	0111	0014	0004	0002	0001	0001	0000	0000	0000	0000
−2,2	0169	0022	0006	0003	0001	0001	0001	0000	0000	0000
−2,6	0247	0032	0009	0004	0002	0001	0001	0000	0000	0000
−3,0	0349	0045	0013	0006	0003	0002	0001	0001	0000	0000
−3,4	0475	0062	0018	0008	0004	0002	0001	0001	0001	0000
−3,8	0630	0082	0024	0010	0005	0003	0002	0001	0001	0001
−4,2	0815	0108	0032	0014	0007	0004	0003	0002	0001	0001
−4,6	1031	0138	0041	0017	0009	0005	0003	0002	0002	0001
−5,0	1279	0173	0052	0022	0011	0006	0004	0003	0002	0001
−5,5	1634	0255	0067	0029	0015	0008	0005	0004	0003	0002
−6,0	2037	0287	0086	0036	0019	0011	0007	0005	0003	0002
−6,5	2485	0360	0108	0046	0023	0014	0009	0006	0004	0003
−7,0	2973	0443	0134	0057	0029	0017	0011	0007	0005	0004
−8,0	4039	0645	0196	0083	0043	0025	0016	0010	0007	0005
−9,0	5265	0897	0276	0117	0050	0035	0022	0015	0010	0008
−10,0	6269	1202	0374	0160	0082	0048	0030	0020	0014	0010
	$\beta_y = 4$									
0,8	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,6	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,4	0001	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,2	0001	0000	0000	0000	0000	0000	0000	0000	0000	0000
0,0	0001	0000	0000	0000	0000	0000	0000	0000	0000	0000
−0,2	0002	0000	0000	0000	0000	0000	0000	0000	0000	0000
−0,4	0003	0000	0000	0000	0000	0000	0000	0000	0000	0000
−0,6	0005	0000	0000	0000	0000	0000	0000	0000	0000	0000
−0,8	0007	0000	0000	0000	0000	0000	0000	0000	0000	0000
−1,0	0010	0001	0000	0000	0000	0000	0000	0000	0000	0000

Окончание таблицы 5

A	C									
	10	20	30	40	50	60	70	80	90	100
$\beta_y = 4$										
−1,4	0018	0001	0000	0000	0000	0000	0000	0000	0000	0000
−1,8	0033	0002	0000	0000	0000	0000	0000	0000	0000	0000
−2,2	0055	0003	0001	0000	0000	0000	0000	0000	0000	0000
−2,6	0088	0006	0001	0000	0000	0000	0000	0000	0000	0000
−3,0	0136	0009	0002	0001	0000	0000	0000	0000	0000	0000
−3,4	0201	0013	0003	0001	0000	0000	0000	0000	0000	0000
−3,8	0289	0019	0004	0001	0000	0000	0000	0000	0000	0000
−4,2	0404	0026	0005	0002	0001	0000	0000	0000	0000	0000
−4,6	0501	0036	0007	0002	0001	0000	0000	0000	0000	0000
−5,0	0732	0048	0010	0003	0001	0001	0000	0000	0000	0000
−5,5	1015	0068	0014	0004	0002	0001	0000	0000	0000	0000
−6,0	1366	0094	0019	0006	0002	0001	0001	0000	0000	0000
−6,5	1788	0126	0025	0008	0003	0002	0001	0000	0000	0000
−7,0	2282	0167	0033	0011	0004	0002	0001	0001	0000	0000
−8,0	3466	0275	0055	0017	0007	0003	0002	0001	0001	0001
−9,0	4838	0429	0087	0027	0011	0005	0003	0002	0001	0001
−10,0	6252	0638	0130	0041	0017	0008	0004	0003	0002	0001

Задача 28. *Прочность элемента конструкции распределена нормально с параметрами* $\mu_x = 100$ *и* $\sigma_x = 50$. *Нагрузка, действующая на элемент, также распределена нормально с параметрами* $\mu_y = 50$ *и* $\sigma_y = 25$. *Вычислить вероятность безотказной работы элемента.*

Имеем $z = -\dfrac{100-50}{50^2+25^2} = -0{,}89442$. Далее, используя аппроксимацию 8 (см. раздел 1.1.1), получаем

$$F(-0{,}89442) = 1 - F(0{,}89442) = 1 - 0{,}852\exp\left\{-\left(\frac{0{,}89442+1{,}5774}{2{,}0637}\right)^{2{,}34}\right\} = 0{,}185337;$$

$$R = 1 - F(-0{,}89442) = F(0{,}89442) = 0{,}814662.$$

Задача 29. *Прочность элемента конструкции распределена логарифмически нормально с медианой* 1000 *ед. и модой* 60 *ед., а нагрузка распределена логарифмически нормально с медианой* 80 *ед. и модой* 40 *ед. Вычислить вероятность безотказной работы элемента.*

Имеем

$$\exp(\mu_{\ln x}) = 100; \quad \exp\left(\mu_{\ln x} - \sigma^2_{\ln x}\right) = 60; \quad \exp(\mu_{\ln y}) = 80; \quad \exp\left(\mu_{\ln y} - \sigma^2_{\ln y}\right) = 40.$$

Тогда

$$\mu_{\ln x} = 4{,}605; \quad \mu_{\ln y} = 4{,}382; \quad \sigma^2_{\ln x} = \mu_{\ln x} - \ln 60 = 0{,}511; \quad \sigma^2_{\ln y} = \mu_{\ln y} - \ln 40 = 0{,}693.$$

Далее

$$z = -\frac{4{,}605 - 4{,}382}{0{,}511 + 0{,}693} = -0{,}18521 \text{ и } R = 1 - F(-0{,}18521) = 1 - 1 + F(0{,}18521) = F(0{,}18521).$$

Применив аппроксимацию из задачи 28, находим

$$R = F(0{,}18521) = 1 - 0{,}852\exp\left\{-\left(\frac{0{,}18521+1{,}5774}{2{,}0637}\right)^{2{,}34}\right\} = 0{,}573257.$$

Задача 30. *Прочность элемента конструкции распределена экспоненциально с* $\lambda_x = 10^{-3}$, *а нагрузка также имеет экспоненциальное распределение с* $\lambda_y = 10^{-4}$. *Вычислить вероятность отказа элемента.*

Вероятность отказа элемента равна

$$1 - R = 1 - \frac{\lambda_y}{\lambda_x + \lambda_y} = \frac{\lambda_x}{\lambda_x + \lambda_y} = 0{,}909.$$

Задача 31. *Прочность элемента конструкции распределена нормально с параметрами* $\mu_x = 200$ *и* $\sigma_x = 30$, *а нагрузка распределена экспоненциально с параметром* $\lambda_y = 10^{-2}$. *Вычислить вероятность безотказной работы элемента.*

Имеем

$$R = 1 - F\left(-\frac{200}{300}\right) - \exp\left\{-\frac{1}{2}(2 \cdot 200 \cdot 10^{-2} - 10^{-4} \cdot 900)\right\}\left\{1 - F\left(-\frac{200 - 10^{-4} \cdot 900}{30}\right)\right\} =$$
$$= 1 - F(-6{,}6667) - \exp(-1{,}955)\{(1 - F(-6{,}6667)\} = F(6{,}6666) - 0{,}14156 \cdot F(6{,}6667).$$

Очевидно, что $F(6{,}6667) \approx 1$, и окончательно получаем $R = 1 - 0{,}14156 = 0{,}858$. Легко видеть, что если поменять местами прочность и нагрузку, то $R = 0{,}142$.

Задача 32. *Прочность элемента конструкции имеет гамма-распределение с параметрами* $\alpha_x = 3$ *и* $\beta_x = 10^{-3}$, *а нагрузка распределена экспоненциально с параметром* $\beta_y = 10^{-3}$. *Вычислить вероятность безотказной работы элемента.*

Имеем $R = 1 - \left(\dfrac{\beta_x}{\beta_x + \beta_y}\right)^2 = 0{,}75$.

Задача 33. *Прочность элемента конструкции распределена по закону Вейбулла с параметрами* $\alpha_y = 800$ *и* $\beta_y = 3$, *а нагрузка распределена нормально с параметрами* $\mu_x = 400$ *и* $\sigma_x = 80$. *Вычислить вероятность безотказной работы элемента.*

Имеем $A = -\dfrac{\mu_x}{\sigma_x} = -\dfrac{400}{80} = -5$; $C = \dfrac{800}{80} = 10$. Из табл. 5 для $\beta = 3$, $A = -5$ и $C = 10$ имеем $1 - R = 0{,}1279$ и $R = 0{,}8721$.

1.1.23. Нецентральное распределение Стьюдента (нецентральное t-распределение)

Описание, применение. Если случайная величина x имеет нормальное распределение со средним δ и единичной дисперсией, а независимая от нее случайная величина χ^2 имеет распределение хи-квадрат с f степенями свободы, то случайная величина $t' = x/\chi^2$ имеет нецентральное t-распределение с f степенями свободы и параметром нецентральности δ.

При $\delta = 0$ нецентральное t-распределение совпадает с центральным (см. раздел 1.1.9). Пусть $\delta(f, t_0, p)$ — значение параметра нецентральности, при котором случайная величина t', имеющая нецентральное t-распределение с параметрами f и δ, превышает значение t_0 с вероятностью p, т. е. $\mathbf{P}\{t'(f, \delta) > t_0\} = p$.

Очевидно, что

$$\delta(f, t_0, p) = -\delta(f, -t_0, p).$$

Таблицы, необходимые для расчетов в случае нецентрального t-распределения, приведены в [24, 29, 96, 97].

При $f > 30$ применимо приближение

$$\delta(f, t_0, p) \approx t_0 - u_{1-p}\left(1 + \frac{t_0^2}{2f}\right)^{\frac{1}{2}},$$

где u_{1-p} — p-квантиль стандартного нормального распределения.

В [97] приведена достаточно простая аппроксимация, основанная на том, то $\operatorname{Arsh}\left[t\sqrt{\frac{3}{2f}}\right]$ имеет нормальное распределение

$$N\left(\operatorname{Arsh}\left[\delta\sqrt{\frac{3}{2f}}\right], 1\right),$$

где $\operatorname{Arsh}(x) = \ln\left(x + \sqrt{x^2+1}\right)$ — гиперболический арксинус, а t — случайная величина, имеющая распределение Стьюдента.

Нецентральное t-распределение используется при планировании эксперимента и оценке мощности критериев проверки гипотез с помощью статистик, основанных на t-критерии.

Задача 34. *Вычислить значение параметра нецентральности δ, при котором случайная величина, имеющая нецентральное t-распределение с $f = 10$ степенями свободы, превысит величину $t_0 = 2$ с вероятностью $p = 0{,}9$.*

Используя нормальную аппроксимацию, имеем

$$\delta(10; 2; 0{,}9) = 2 - u_{0,1}\left(1 + \frac{2^2}{2 \cdot 10}\right) = 2 - u_{0,1} \cdot 1{,}095445.$$

Пользуясь аппроксимацией 15 из раздела 1.1.1, имеем

$$u_{0,1} = 4{,}91 \cdot \left(0{,}1^{0,14} - 0{,}9^{0,14}\right) = -1{,}2812615 \quad \text{и} \quad \delta = 2 + 1{,}2812615 \cdot 1{,}095445 = 3{,}40356.$$

1.1.24. Нецентральное распределение Пирсона (нецентральное распределение хи-квадрат)

Описание, применение. Сумма квадратов n независимых нормально распределенных случайных величин с единичной дисперсией, среди которых по крайней мере одна имеет ненулевое математическое ожидание, подчиняется нецентральному χ'^2-распределению.

Нецентральному χ'^2-распределению подчинена сумма

$$\chi'^2 = \sum_{i=1}^{n} \frac{1}{\sigma^2}(x_i - \overline{x} + \overline{\gamma})^2,$$

где x_i — независимые нормально распределенные величины с нулевым средним и общей дисперсией σ^2 ; γ_i — постоянные со средним $\overline{\gamma} = \frac{1}{n}\sum_{i=1}^{n}\gamma_i$ и $\overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$.

Величина χ'^2 подчиняется нецентральному χ^2-распределению с $f = n$ степенями свободы и параметром нецентральности $a = \sum_{i=1}^{n}\frac{1}{\sigma^2}(\gamma_i - \overline{\gamma})^2$.

Таблицы нецентрального χ'^2-распределения приведены в [24].

При $a \to 0$ $\quad \chi'^2(f, a) \to \chi^2(f)$.

При $a \to \infty$ $\quad \dfrac{\chi'^2(f, a) - n - a}{\sqrt{2(n + 2a)}} \to N(0{,}1)$.

Пирсон предложил преобразование, позволяющее вычислить нецентральное χ^2-распределение через центральное [98]:

$$\chi^2(f') = \frac{f+2a}{f+3a}\left\{\chi^2(f,a) + \frac{a^2}{f+3a}\right\},$$

где $\chi^2(f')$ — центральное χ^2-распределение с $f' = \dfrac{(n+2a)^3}{(n+3a)^2}$ степенями свободы.

При $a \to 0$ $\chi^2(f')$ отличается от функции точного распределения хи-квадрат на величину порядка a^2, а при $a \to \infty$ на величину порядка $\dfrac{1}{a}$.

Более простое, но менее точное преобразование предложено Патнайком [99]: $\chi''^2(f'') = \dfrac{f+a}{f+2a}\chi^2(f,a)$, где $f'' = \dfrac{(f+9)^2}{f+2a}$. Полезно следующее преобразование [97]: обычная величина $c\chi^2(f')$ имеет то же распределение, что и нецентральная $\chi'^2(f,a)$, если $c = \dfrac{f+2a}{f+a}$ и $f' = \dfrac{f+a^2}{f+2a}$.

Нецентральное χ^2-распределение применяется при планировании эксперимента для проверки гипотез, когда используются статистики, основанные на случайных величинах χ^2.

Задача 35. *Вычислить 95%-ю квантиль случайной величины, имеющей нецентральное χ^2-распределение с $f = 30$ степенями свободы и параметром нецентральности $a = 10$.*

Из соотношения

$$\chi^2(f') = \frac{f+2a}{f+3a}\left\{\chi^2(f,a) + \frac{a^2}{f+3a}\right\}$$

следует $\chi^2(f,a) = \chi^2(f')\dfrac{f+3a}{f+2a} - \dfrac{a^2}{f+3a}$, или

$$\chi^2(30,10) = \chi^2(f')\cdot\frac{30+30}{30+20} - \frac{100}{30+30} = 1{,}2\chi^2(f') - 1{,}6667$$

$$\text{и} \quad f' = \frac{(f+2a)^3}{(f+3a)^2} = \frac{50^2}{60^2} = 34{,}722.$$

Найдем теперь верхнюю 95%-ю квантиль обычного χ^2-распределения с $f = 34{,}722$ степенями свободы. Воспользуемся аппроксимацией 10 (Хоглина) из раздела 1.1.8:

$$\chi_p^2 = \left\{1{,}06807\sqrt{f} + 2{,}13161\left[-\lg(1-p)\right]^{\frac{1}{2}} - 0{,}04589\sqrt{f}\cdot\left[-\lg(1-p)\right]^{\frac{1}{2}} - 1{,}37266\right\}^2 =$$

$$= \left\{1{,}06807\sqrt{34{,}722} + 2{,}13161\cdot 1{,}140627 - 0{,}04589\sqrt{34{,}722}\cdot 1{,}140627 - 1{,}37266\right\}^2 = 49{,}6168.$$

1.1.25. Нецентральное распределение Фишера (нецентральное F-распределение)

Описание, применение. Если $\chi_1'^2$ — случайная величина, имеющая нецентральное распределение хи-квадрат с f_1 степенями свободы и параметром нецентральности a, а χ_2^2 — независимая от нее случайная величина, подчиняющаяся распределению хи-квадрат с f_2 степенями свободы, то случайная величина $F^* = \dfrac{f_2\chi_1'^2}{f_1\chi_2^2}$ имеет нецентральное F-распределение.

Для проведения расчетов, связанных с нецентральным F-распределением, могут быть использованы графики, приведенные в [29, 100]. Нецентральное F-распределение используется при планировании эксперимента и оценки мощности критериев проверки гипотез, основанных на применении F-распределения.

Табулирование нецентрального F-распределения затруднено из-за большого количества независимых переменных (f_1, f_2, a), поэтому разработаны его аппроксимации обычным (центральным) F-распределением [101, 102]. Наиболее проста и точна аппроксимация М. Тику [101], в соответствии с которой величина $F = \dfrac{F^* + c}{h}$ имеет центральное F-распределение с параметром b и f_2 степенями свободы, где

$$b = \frac{f_2 - 2}{2}\left(\sqrt{\frac{E}{E-4}} - 1\right); \quad h = \frac{b}{f_1}\frac{1}{2b + f_2 - 2}\frac{H}{K}; \quad E = \frac{H^2}{K^3};$$

$$c = \frac{f_2}{f_2 - 2}\left(h - \frac{f_1 + a}{f_1}\right); \quad K = (f_1 + a)^2 + (f_2 - 2)(f_1 + 2a);$$

$$H = 2(f_1 + a)^3 + 3(f_1 + a)(f_1 + 2a)(f_2 - 2) + (f_1 + 3a)(f_2 - 2)^2.$$

В [97] предложена упрощенная формула для нецентрального F-распределения. Случайная величина $F^*_{f_1, f_2}(a)$, имеющая нецентральное F-распределение со степенями свободы f_1 и f_2 и параметром нецентральности a, распределена как величина $\left(1 + \dfrac{a}{f_1}\right)F(f_1^*, f_2)$, где $f_1^* = \dfrac{(f_1 + a)^2}{f_1 + 2a}$ и $F(\nu_1, \nu_2)$ — обычная F-величина с ν_1 и ν_2 степенями свободы.

Более точные аппроксимации можно найти в [115].

Задача 36. *Вычислить верхнюю* 10%*-ю точку нецентрального* F'*-распределения с параметрами* $f_1 = 10$, $f_2 = 12$ *и* $a = 2$.

Используем аппроксимацию Тику. Имеем

$$H = 2 \cdot (10 + 2)^3 + 3 \cdot (10 + 2) \cdot (10 + 4) \cdot (12 - 2) + (10 + 6) \cdot (12 - 2)^2 = 10096;$$

$$K = (10 + 2)^2 + (12 - 2) \cdot (10 + 4) = 284; \quad E = \frac{10096^2}{284^3} = 4{,}4498325;$$

$$h = \frac{10{,}72558}{10}\frac{1}{2 \cdot 10{,}72558 + 12 - 2} \cdot \frac{10096}{284} = 1{,}212314;$$

$$b = \frac{12 - 2}{2} \cdot \left(\sqrt{\frac{4{,}4498325}{0{,}4498325}} - 1\right) = 10{,}72558; \quad c = \frac{12}{12 - 2} \cdot \left(1{,}212314 - \frac{10 + 2}{10}\right) = 0{,}0147768.$$

Исходя из того, что случайная величина $F = \dfrac{F^* + c}{h}$ имеет центральное F-распределение, находим искомую величину, где F — обычная F-величина с $b = 10{,}725933$ и $f_2 = 12$ степенями свободы. Для ее вычисления воспользуемся аппроксимацией 9 из раздела 1.1.10. Для $p = 0{,}9$ (верхняя 10%-я точка; не путать h и b в этих обозначениях с обозначениями, ранее применяемыми) имеем

$$a = 1{,}1131; \quad b = 0{,}77; \quad c = 0{,}527;$$

$$h = \frac{2f_1 f_2}{f_1 + f_2} = \frac{2 \cdot 10{,}72558 \cdot 12}{10{,}72558 + 12} = 11{,}32705612; \quad g = \frac{12 - 10{,}72558}{10{,}72558 \cdot 12} = 9{,}901791 \cdot 10^{-3}.$$

Тогда

$$\lg F_{0,9}(10{,}72558; 12) = 1{,}1131 \cdot (11{,}32705612 - 0{,}77)^{\frac{1}{2}} - 0{,}527 \cdot 9{,}901791 \cdot 10^{-3} = 0{,}3373623$$

$$\text{и} \quad F_{0,9}(0{,}72558; 12) = 2{,}1745; \quad F^*_{0,9} = 1{,}2123267 \cdot 2{,}1745 - 0{,}014792021 = 2{,}621412.$$

Используем теперь второе приближение. Находим $f_1^* = \dfrac{(10+2)^2}{10+4} = 10{,}285714$ и вычисляем $f_{0,9}(10,285714; 12)$ по аналогии ($a = 1{,}1131$; $b = 0{,}77$; $c = 0{,}527$)

$$b = \frac{2 \cdot 10{,}285714 \cdot 12}{10{,}285714 + 12} = 11{,}0769229; \quad g = \frac{12 - 10{,}285714}{10{,}285714 \cdot 12} = 0{,}01388889;$$

$$\lg_{10} F_{0,9}(10{,}285714; 12) = 1{,}1131 \cdot (11{,}0769229 - 0{,}77)^{\frac{1}{2}} - 0{,}527 \cdot 0{,}01388889 = 0{,}339393;$$

$$F_{0,9}(10{,}285714; 12) = 2{,}184707 \quad \text{и} \quad F_{0,9}^*(10; 12; 2) = \left(1 + \frac{2}{10}\right) \cdot 2{,}184707 = 2{,}621648.$$

Видно, что погрешность аппроксимации вторым способом удовлетворительна, а сам способ много проще в вычислительном отношении.

1.2. Дискретные распределения

1.2.1. Биномиальное распределение (распределение Бернулли)

Свойства, применение. Если событие осуществляется в единичном испытании с некоторой постоянной вероятностью, то число появлений события в последовательности независимых испытаний подчинено распределению Бернулли. При испытаниях невосстанавливаемых изделий на фиксированную наработку число отказов подчиняется биномиальному распределению. Оно широко применяется при выборочном приемном контроле качества продукции и статистическом предупредительном контроле технологических процессов в производстве.

Свойства

Распределение вероятностей

$$f(x;n,p) = C_n^x p^x (1-p)^{n-x}, \quad 0 \leqslant p \leqslant 1$$

— вероятность появления события ровно x раз в серии из n испытаний, при условии, что в единичном испытании вероятность его появления равна p.
Функция распределения

$$F(x;n,p) = \sum_{i=0}^{x} C_n^i \, p^i (1-p)^{n-i}$$

— вероятность появления событий $\leqslant x$ раз в серии из n испытаний.

Среднее	$\mathbf{M}(x) = np$
Дисперсия	$\mathbf{D}(x) = np(1-p)$
Коэффициент вариации	$v = \left(\dfrac{1-p}{np}\right)^{\frac{1}{2}}$
Коэффициент асимметрии	$\alpha_3 = (1-2p)[np(1-p)]^{-\frac{1}{2}}$
Коэффициент эксцесса	$\alpha_4 = 3 - \dfrac{6}{n} + \dfrac{1}{np(1-p)}$

Таблицы вероятностей биномиального распределения приведены в [16, 23, 29, 44]. Однако появление современных компактных и весьма мощных микрокалькуляторов позволяют достаточно быстро производить расчеты без применения таблиц [30, 31, 33].

При $n \to \infty$, $p \to 0$ и $np = \mathrm{const}$ биномиальное распределение сводится к распределению Пуассона (см. раздел 1.2.2) с параметром $\lambda = np$ (приближение приемлемо при $n > 10$ и $p < 0{,}1$). В силу взаимосвязи распределения Пуассона с распределением хи-квадрат (см. раздел 1.1.8) функция биномиального распределения может быть выражена через интеграл вероятностей χ^2-распределения:

$$F(x;n,p) = \sum_{i=0}^{x} \frac{(np)^i}{i!}\, e^{-np} = P_{\chi^2}(2np, 2x+2),$$

где $P_{\chi^2}(y,\nu) = 1 - F\chi^2_\nu(x)$ — интеграл вероятностей χ^2-распределения с ν степенями свободы, $F_{\chi^2_\nu}(x)$ — функция распределения χ^2_ν с ν степенями свободы.

При $n \to \infty$ биномиальное распределение стремится к нормальному со средним $\mu = np$ и дисперсией $\sigma^2 = np(1-p)$. Сходимость удовлетворительна при $np(1-p) > 5$ и $0{,}1 \leqslant p \leqslant 0{,}9$ или при $np(1-p) > 25$ и любом p. Наилучшая сходимость обеспечивается при $p = 0{,}5$ и ухудшается при $p < \dfrac{1}{n+1}$ и $p > \dfrac{1}{n+1}$.

Таким образом, в указанных диапазонах имеет место

$$F(x;n,p) = \Phi\left(\frac{x - np + \dfrac{1}{2}}{\sqrt{np(1-p)}}\right),$$

где $\Phi(\ldots)$ — функция стандартного нормального распределения.

Биномиальное распределение может быть выражено через функцию бета-распределения (см. раздел 1.1.7)

$$F(x;n,p) = 1 - I_p(x1, n-x) = I_{1-p}(n-x, x+1),$$

где $I_p(a,b)$ — неполная бета-функция, и через функцию F-распределения (см. раздел 1.1.10)

$$F(x;n,p) = G\left[\frac{x+1}{n-x}\frac{1-p}{p}; 2(n-x), 2(x+1)\right],$$

где $G(y; f_1, f_2)$ — функция F-распределения относительно переменной y с f_1 и f_2 степенями свободы.

Приведенные соотношения позволяют использовать таблицы и аппроксимации бета- и F-распределений для расчетов при биномиальном распределении.

Пуассоновские и нормальные приближения, рассмотренные выше, относительно грубы. Поэтому был предложен ряд точных приближений, например, в [104]:

$$F(x;n,p) = P_{\chi^2}(2y, 2x+2),$$

где при $y = y^* = \dfrac{p(2n-x)}{2-p}$ погрешность $\approx \dfrac{1}{n^2}$ при $p < 0{,}2$ и $n \leqslant 300$ [17], а при $y = y^{**} = y^*\left\{1 + [x(x+2) + xy^* - 2y^*]^2\left[6(2n-x^2)^{-1}\right]\right\}^{-1}$ погрешность $\approx \dfrac{1}{n^4}$ (при $p \leqslant 0{,}12$ и $n \leqslant 300$ [17]).

Моленаром [55] предложены весьма точные аппроксимации биномиального распределения нормальным:

— для „хвостов“ распределения ($0{,}005 \leqslant p \leqslant 0{,}05$ и $0{,}93 \leqslant p \leqslant 0{,}995$)

$$F(x;n,p) = \Phi\left\{(4x+1)(1-p)^{\frac{1}{2}} - [(4n-4x-1)p]^{\frac{1}{2}}\right\};$$

— для распределения между „хвостами“

$$F(x;n,p) = \Phi\left\{(4x+2{,}5)(1-p)^{\frac{1}{2}} - [(4n-4x-1)p]^{\frac{1}{2}}\right\};$$

при этом должны выполняться приблизительно следующие соотношения между n и p:

$$n = 3; \quad 0{,}25 \leqslant p \leqslant 0{,}75;$$
$$n = 30; \quad 0{,}40 \leqslant p \leqslant 0{,}60;$$
$$n = 300; \quad 0{,}46 \leqslant p \leqslant 0{,}54.$$

Для более точной аппроксимации используется формула [105]

$$F(x;n,p) = \Phi\left\{[A(x,\lambda)]^{\frac{1}{2}} - [A(n-x-1, -\lambda)]^{\frac{1}{2}}\right\},$$

где

$$A(x,\lambda) = 2(x+1)(1+\lambda) - \frac{15 + 12\lambda + 5\lambda^2}{18} + \frac{(3 + 6\lambda + 7\lambda^2)(2x + 1 + n\lambda)}{36n(1-\lambda^2)}, \quad \lambda = 1 - p.$$

При табулировании биномиального распределения возникают трудности из-за того, что его параметр p одновременно входит в выражения для среднего и для дисперсии. Специальные преобразования исходной биномиально распределенной случайной величины позволяют устранить эти трудности. В [106] показано, что случайная величина $z = 2\arcsin\sqrt{x/n}$, где x — биномиально распределенная случайная величина, имеет при $n > 50$ нормальное распределение с математическим ожиданием $\mu = 2\arcsin\sqrt{p}$ и дисперсией $\sigma^2 = 1/n$. При умеренных n ($n < 50$) Анскомбом [107] предложено уточненное преобразование

$$z^* = \arcsin\sqrt{\frac{x+\frac{3}{8}}{n+\frac{3}{4}}},$$

удовлетворительно аппроксимирующееся нормальным распределением с $\mu = \arcsin\sqrt{\dfrac{p+\frac{3}{8n}}{1+\frac{3}{4n}}}$ и $\sigma^2 = \dfrac{1}{4n+2}$.

Характеристическим свойством биномиального распределения — дисперсия меньше среднего — можно руководствоваться при выборе между ним и отрицательно-биномиальным (см. раздел 1.2.3), для которого дисперсия больше среднего, и распределением Пуассона (см. раздел 1.2.2), у которого дисперсия равна среднему.

Задача 37. *Вероятность появления дефектного изделия в производстве равна* 0,1. *Вычислить вероятность появления в партии из* 60 *изделий не более* 10 *дефектных.*

Имеем $p = 0{,}1$; $n = 60$; $x = 10$ и $np = 6$. Необходимо вычислить величину $F(10; 0,1; 60)$. Непосредственный расчет (точное решение) дает

$$F(10;0{,}1;60) = \sum_{i=0}^{10} C_{60}^i \cdot 0{,}1^i \cdot 0{,}9^{60-i} = C_{60}^0 \cdot 0{,}9^{60} + C_{60}^1 \cdot 0{,}1 \cdot 0{,}9^{59} + C_{60}^2 \cdot 0{,}1^2 \cdot 0{,}9^{58} =$$
$$= C_{60}^3 \cdot 0{,}1^3 \cdot 0{,}9^{57} + C_{60}^4 \cdot 0{,}1^4 \cdot 0{,}9^{56} + C_{60}^5 \cdot 0{,}1^5 \cdot 0{,}9^{55} + C_{60}^6 \cdot 0{,}1^6 \cdot 0{,}9^{54} + C_{60}^7 \cdot 0{,}1^7 \cdot 0{,}9^{53} +$$
$$+ C_{60}^8 \cdot 0{,}1^8 \cdot 0{,}9^{52} + C_{60}^9 \cdot 0{,}1^9 \cdot 0{,}9^{51} + C_{60}^{10} \cdot 0{,}1^{10} \cdot 0{,}9^{50} = 0{,}96570865.$$

Аппроксимация с помощью распределения Пуассона:

$$F(10;60;0{,}1) = \sum_{i=0}^{10} \frac{6^i}{i!} \cdot e^{-6} = 2{,}478752 \cdot 10^{-3} \times$$
$$\times \left(1 + 6 + \frac{36}{2} + \frac{216}{6} + \frac{1296}{24} + \frac{7776}{120} + \frac{46656}{720} + \frac{279936}{5040} + \frac{1679616}{40320} + \frac{10077696}{362880} + \frac{60466176}{3628800}\right) =$$
$$= 0{,}9573799 \quad (\delta = 0{,}87\%).$$

Если воспользоваться соотношением между распределением Пуассона и χ^2-распределением $\sum\limits_{i=0}^{10} \dfrac{6^i}{i!} e^{-6} = P_{\chi^2}(2 \cdot 6; 20+2) = P_{\chi^2}(12{,}22)$ и применить аппроксимацию 3 (см. раздел 1.1.8), то получим из соотношения $\chi_p^2 = f\left(1 - \dfrac{2}{9f} + u_p\sqrt{\dfrac{2}{9f}}\right)^3$, имея в виду, что $\chi_3^2 = 12$ и $f = 22$, уравнение $12 = 22\left(1 - \dfrac{2}{9 \cdot 22} + u_p\sqrt{\dfrac{2}{9 \cdot 22}}\right)^3$, решением которого является величина нормальной квантили $u_p = 1{,}719747$.

Используем аппроксимацию 8 из раздела 1.1.1:

$$F(u_p) = 1 - 0{,}852\exp\left[-\left(\frac{1{,}719747 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right] = 0{,}957302.$$

Полученное значение вероятности $P_{\chi^2}(12{,}22) = 1 - F(\chi^2 < 12) \approx 0{,}957302$ очень близко к точной величине 0,95778.

Используем теперь нормальное приближение

$$F(x; n, p) = \Phi\left(\frac{x - np + \dfrac{1}{2}}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{10 - 6 + 0{,}5}{\sqrt{6 \cdot 0{,}9}}\right) = \Phi(1{,}93649).$$

С помощью аппроксимации 18 из раздела 1.1.1 получаем

$$\Phi(1{,}93649) = 1 - \frac{1}{2}\exp(-0{,}717 \cdot 1{,}93649 - 0{,}416 \cdot 1{,}93649^2) = 0{,}97378 \quad (\delta = 0{,}8\%).$$

С помощью приближения бета-распределением имеем

$$F(x; n, p) = G\left(\frac{11}{50} \cdot \frac{0{,}9}{0{,}1}; 100{,}22\right) = G(1{,}98; 100{,}22).$$

Применим аппроксимацию 4 (см. раздел 1.1.10):

$$u_p = \frac{1{,}98^{\frac{1}{3}} \cdot \left(1 - \dfrac{2}{9 \cdot 22}\right) - \left(1 - \dfrac{2}{9 \cdot 100}\right)}{\left(\dfrac{2}{9 \cdot 100} + \dfrac{2}{9 \cdot 22} \cdot 1{,}98^{\frac{2}{3}}\right)^{\frac{1}{2}}} = 1{,}820406.$$

Далее, используя аппроксимацию 8 из раздела 1.1.1, получаем:

$$F(u_p) = F(1{,}820406) = 1 - 0{,}852\exp\left[-\left(\frac{1{,}820406 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right] = 0{,}96567,$$

что очень близко к точному значению 0,96579.

Рассмотрим *нормализующие преобразования*

$$z = 2\arcsin\left(\frac{10}{60}\right)^{\frac{1}{2}} = 0{,}84106867; \quad \mu = 2\arcsin\sqrt{0{,}1} = 0{,}643501;$$

$$\sigma = \frac{1}{60} = 0{,}1290994; \quad u_p = \frac{0{,}84106867 - 0{,}643501}{0{,}1290994} = 1{,}5303518.$$

Далее

$$F(u_p) = 1 - 0{,}852\exp\left[-\left(\frac{1{,}5303518 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right] = 0{,}9371266.$$

Для уточненного преобразования имеем

$$z^* = \arcsin\sqrt{\frac{10 + \dfrac{3}{8}}{60 + \dfrac{3}{4}}} = 0{,}426028; \quad \mu = \arcsin\sqrt{\frac{0{,}1 + \dfrac{3}{8 \cdot 60}}{1 + \dfrac{3}{4 \cdot 60}}} = 0{,}3298829;$$

$$\sigma = \frac{1}{\sqrt{4 \cdot 60} + 2} = 0{,}0642824; \quad u_p = \frac{0{,}426028 - 0{,}3298929}{0{,}0642824} = 1{,}49551.$$

Окончательно

$$F(1{,}49551) = 1 - 0{,}852\exp\left[-\left(\frac{1{,}49551 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right] = 0{,}93271.$$

Задача 38. *Имеем* $p = 0{,}5$ *и* $n = 5$. *Вычислить* $F(3; 0,5, 5)$, *используя аппроксимацию Моленара.*

Имеем $4x + 2{,}5 = 1{,}45$; $4n - 4x - 1 = 20 - 12 - 1 = 7$ и

$$F(3; 0{,}5, 5) = \Phi\{\sqrt{14{,}5 \cdot 0{,}5} - \sqrt{7 \cdot 0{,}5}\} = \Phi(0{,}8217537) = 0{,}793382.$$

Для сравнения найдем точное значение:

$$F(3; 0{,}5, 5) = \sum_{i=0}^{3} c_5^i \cdot 0{,}5^i \cdot 0{,}5^{5-i} = 0{,}5^5 \cdot (1 + 5 + 10 + 10) = 0{,}8125 \quad (\delta = 2{,}3\%).$$

1.2.2. Распределение Пуассона

Описание, применение. Если вероятность появления независимых событий в малом промежутке времени Δt пропорциональна Δt, то число их появлений имеет распределение Пуассона. Распределение Пуассона широко применяется в теории массового обслуживания. Часто оно используется для аппроксимации биномиального, так как легче табулируется (имеет только один параметр). Распределению Пуассона подчинено число отказов невосстанавливаемых электронных приборов в течение периода приработки. Если наработка на отказ изделия является случайной величиной, распределенной экспоненциально с параметром λ, то число отказов в интервале времени t подчиняется распределению Пуассона с параметром $t\lambda$.

Свойства

Распределение вероятностей

$$f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots; \quad \lambda > 0;$$

$f(x; \lambda) = \max$ при $x = [\lambda]$ (наибольшее целое число $\leqslant \lambda$)

Функция распределения	$F(x; \lambda) = \sum_{i=0}^{x} \frac{\lambda^i e^{-\lambda}}{i!}$
Среднее	$\mathbf{M}(x) = \lambda$
Дисперсия	$\mathbf{D}(x) = \lambda$
Коэффициент вариации	$v = \lambda^{-\frac{1}{2}}$
Коэффициент асимметрии	$\alpha_3 = \lambda^{-\frac{1}{2}}$
Коэффициент эксцесса	$\alpha_4 = 3 + \frac{1}{\lambda}$

При $\lambda > 9$ распределение Пуассона можно аппроксимировать нормальным с $\mu = \lambda$ и $\sigma = \sqrt{\lambda}$, т. е. $\Phi\left(\frac{x + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) = F(x; \lambda)$ (1/2 — поправка Йэтса на непрерывность).

Распределение Пуассона является предельной формой биномиального распределения (см. раздел 1.2.1), т. е.

$$\lim_{\substack{n \to \infty \\ np \to \lambda}} \sum_{i=0}^{x} C_n^i p^i (1-p)^{n-i} = \frac{\lambda^x}{x!} e^{-\lambda}.$$

Функция распределения Пуассона может быть выражена через функцию χ^2-распределения

$$\sum_{i=0}^{x} \frac{\lambda^i e^{-\lambda}}{i!} = 1 - P(2\lambda, 2x+2),$$

где $P(2\lambda, 2x + 2)$ — функция χ^2-распределения относительно переменной 2λ с $f = 2(x+1)$ степенями свободы.

По аналогии с биномиальным распределением Анскомб [107] показал, что случайная величина $2\sqrt{x}$ при $\lambda \to \infty$ распределена асимптотически нормально с $\mu = 2\sqrt{\lambda}$ и $\sigma = 1$.

С учетом поправки на непрерывность можно для прикидочных расчетов использовать формулы

$$\sum_{i=0}^{x} \frac{\lambda^i e^{-\lambda}}{i!} = \begin{cases} \Phi(2\sqrt{x+1} - 2\lambda) & \text{— на „хвостах“ распределения;} \\ \Phi(2\sqrt{x+0{,}75} - 2\lambda) & \text{— между „хвостами“ распределения.} \end{cases}$$

Высокой точностью обладает аппроксимация

$$\sum_{i=0}^{x} \frac{\lambda^i e^{-\lambda}}{i!} = \Phi\left[2\Big(x + \frac{\gamma+4}{9}\Big)^{\frac{1}{2}} - 2\Big(\lambda + \frac{\gamma-8}{36}\Big)^{\frac{1}{2}}\right], \quad \text{где} \quad \gamma = \Big(x = \lambda + \frac{1}{6}\Big)^2 \lambda^{-1}$$

(здесь $\Phi(\ldots)$ — функция стандартного нормального распределения, обозначенная в разделе 1.1.1 как $F(z)$).

Учитывая связь между распределением Пуассона и χ^2-распределением, для его аппроксимации можно использовать аппроксимации для распределения хи-квадрат.

Сумма случайных величин, распределенных по закону Пуассона, также будет иметь распределение Пуассона с параметром, равным сумме параметров слагаемых. Существуют обширные таблицы распределения Пуассона, например, в [23–25, 57].

Характеристической особенностью распределения Пуассона является равенство среднего и дисперсии.

Задача 39. *Случайная величина имеет распределение Пуассона с параметром* $\lambda = 5$. *Вычислить вероятность того, что случайная величина не превысит значение* 3.

Прямой расчет по формуле распределения:

$$F(3;5) = \sum_{i=0}^{3} \frac{5^i e^{-5}}{i!} = e^{-5} \cdot (1 + 5 + 12{,}5 + 20{,}833) = 0{,}265023669.$$

Грубая нормальная аппроксимация (см. аппроксимацию для $\Phi(x)$ в разделе 1.1.1):

$$F(3;5) = \Phi\left[\frac{3 + \frac{1}{2} - 5}{\sqrt{5}}\right] = \Phi(-0{,}6708) = 1 - 0{,}852\exp\left\{-\left(\frac{-0{,}6708 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\} = 0{,}26367.$$

Нормальная аппроксимация между „хвостами“:

$$F(3;5) = \Phi\{2\sqrt{3+0{,}75} - 2\sqrt{5}\} = \Phi(-0{,}5991526) =$$

$$= 1 - 0{,}852\exp\left\{-\left(\frac{-0{,}5991526 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\} = 0{,}2843.$$

Уточненная нормальная аппроксимация:

$$\gamma = \frac{\left(3 - 5 + \frac{1}{6}\right)^2}{5} = 0{,}67222; \quad \left(x + \frac{\gamma + 4}{9}\right)^{\frac{1}{2}} = 1{,}8759359; \quad \left(\lambda + \frac{\gamma - 8}{36}\right)^{\frac{1}{2}} = 2{,}19008;$$

Далее

$$F(3;5) = \Phi(2 \cdot 1{,}8759359 - 2 \cdot 2{,}19008) = \Phi(-0{,}6282882) =$$
$$= 1 - 0{,}852 \exp\left\{-\left(\frac{-0{,}6282882 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\} = 0{,}27573 \quad (\delta = 3{,}8\%).$$

Погрешность аппроксимации достаточно высока, даже с учетом примененной простой аппроксимации для стандартного нормального распределения.

1.2.3. Отрицательное биномиальное распределение

Описание, применение. Если p — вероятность появления события в единичном испытании, то случайное число x неудачных испытаний до появления m-го успеха подчиняется отрицательному биномиальному распределению. Распределение используется при планировании запуска изделий в производство для получения требуемого количества годных изделий при известном проценте выхода годных, при планировании объема испытаний до получения заданного числа отказов.

Свойства

Распределение вероятностей

$$f(x;m,p) = C^m_{x+m-1} p^m (1-p)^x, \quad m = 0, 1, 2, \ldots; \quad x = 0, 1, 2, \ldots$$

Функция распределения $\quad F(x;m,p) = \sum_{i=1}^{m} C^m_{m+i-1} p^m (1-p)^i$

Среднее $\quad \mathbf{M}(x) = \dfrac{m(1-p)}{p}$

Дисперсия $\quad \mathbf{D}(x) = \dfrac{m(1-p)}{p^2}$

Коэффициент вариации $\quad v = \dfrac{1}{\sqrt{m(1-p)}}$

Коэффициент асимметрии $\quad \alpha_3 = (2-p)[m(1-p)]^{-\frac{1}{2}}$

Коэффициент эксцесса $\quad \alpha_4 = 3 + \dfrac{6}{m} + \dfrac{p^2}{m(1-p)}$

В отличие от биномиального (см. раздел 1.2.1), при отрицательном биномиальном распределении множество возможных значений случайной величины не ограничено сверху. Функция распределения отрицательного биномиального распределения может быть вычислена с помощью таблиц для биномиального распределения [108].

Известны аппроксимации отрицательного биномиального распределения положительным биномиальным [109] и гамма-распределением [110]. Характеристической особенностью отрицательного биномиального распределения является то, что его дисперсия больше среднего (у обычного биномиального — наоборот).

Задача 40. *Вероятность получения дефектного изделия равна* 0,1. *Вычислить вероятность того, что будут произведены* 50 *годных изделий до появления* 10*-го дефектного изделия* ($m = 10$; $p = 0{,}1$; $x = 50$). *Вычислить вероятность того, что до появления* 2*-го дефектного изделия будут произведены не менее* 5 *годных изделий.*

Вычисляем $f(x; m,p) = C_{50+10-1}^{10} \cdot 0{,}1^{10} \cdot 0{,}9^{50} = C_{59}^{1} \cdot 0{,}1^{10} \cdot 0{,}9^{50} = 0{,}03238$ — такова вероятность того, что потребуется произвести 50 изделий до появления 10-го дефектного изделия.

Вероятность того, что до появления 2-го дефектного изделия будет произведено не более 5 годных изделий, равна

$$F(x; m,p) = \sum_{\imath=1}^{5} C_{2+\imath-1}^{\imath} \cdot 0{,}1^2 \cdot 0{,}9^{\imath} =$$
$$= 0{,}1^2 \cdot (C_2^1 \cdot 0{,}9 + C_3^2 \cdot 0{,}9^2 + \ldots + C_5^4 \cdot 0{,}9^4 + C_6^5 \cdot 0{,}9^5) = 0{,}13969.$$

1.2.4. Распределение Паскаля

Описание, применение. Если в схеме, рассмотренной выше (см. раздел 1.2.3) для отрицательного биномиального распределения, в качестве случайной величины принять число удачных испытаний до появления m-го успеха (включая и этот успех), то она будет подчинена распределению Паскаля.

Свойства

Распределение вероятностей

$$f(x; m, p) = C_{x-1}^{m-1} p^m (1-p)^{x-m}, \quad x = m, m+1, \ldots$$

— это вероятность того,что в x испытаниях наблюдаемое событие произойдет равно m раз, если вероятность появления его в каждом испытании равна p.
Функция распределения

$$F(x; m, p) = \sum_{i=m}^{x} C_{i-1}^{m} p^i (1-p)^{i-m}$$

— это вероятность появления m успехов не более, чем за x испытаний.

Среднее $\mathbf{M}(x) = \dfrac{m}{p}$

Дисперсия $\mathbf{D}(x) = \dfrac{m(1-p)}{p^2}$

Коэффициент вариации $v = \left(\dfrac{1-p}{m}\right)^{\frac{1}{2}}$

Задача 41. *Если вероятность отказа изделия в одном испытании равна* 0,1, *то какова вероятность того, что понадобятся* 50 *испытаний до появления* 5 *отказов* ($p = 0{,}1$; $m = 5$; $x = 50$)? *Вычислить вероятность того, что не более, чем за* 8 *испытаний, будут зафиксированы* 2 *отказа* ($x = 50$; $m = 2$; $p = 0{,}1$).

Вероятность того, что понадобятся 50 испытаний до появления 5 отказов:

$$f(x; 5; 0{,}1) = C_{49}^4 \cdot 0{,}1^5 \cdot 0{,}9^{45} = 211876 \cdot 10^{-5} \cdot 8{,}72796 \cdot 10^{-3} = 0{,}01849.$$

Вероятность того, что не более, чем за 8 испытаний, будут зафиксированы 2 отказа, равна

$$F(x; m,p) = \sum_{\imath=2}^{8} C_{\imath-1}^{2} p^{\imath} (1-p)^{\imath-2} = \sum_{\imath=2}^{8} C_{i-1}^{2} \cdot 0{,}9^{\imath} \cdot 0{,}9^{\imath-2} = 0{,}011224.$$

1.2.5. Геометрическое распределение (распределение Фарри)

Описание, применение. Если p — вероятность появления события в одном испытании, то число испытаний до появления события подчинено геометрическому распределению, с помощью которого можно определить объем выборки, необходимой для получения одного отказа по заданной вероятности отказа одного прибора.

Свойства

Распределение вероятностей	$f(x;p) = p(1-p)^{x-1}, \quad x = 1, 2, \ldots; \quad 0 \leqslant p \leqslant 1$
Функция распределения	$F(x;p) = \sum\limits_{i=1}^{x} p(1-p)^{i-1} = 1-(1-p)^x$
Среднее	$\mathbf{M}(x) = \dfrac{1}{p}$
Дисперсия	$\mathbf{D}(x) = \dfrac{1-p}{p^2}$
Коэффициент вариации	$v = \sqrt{1-p}$
Коэффициент асимметрии	$\alpha_3 = \dfrac{2-p}{\sqrt{1-p}}$
Коэффициент эксцесса	$\alpha_4 = 9 + \dfrac{p^2}{1-p}$

Очевидно, что геометрическое распределение следует из распределения Паскаля (см. раздел 1.2.4) при $m = 1$. При $p \to 0$ геометрическое распределение переходит в экспоненциальное (см. раздел 1.1.4).

Задача 42. *Вероятность безотказной работы изделия равна* 0,95. *Вычислить вероятность того, что для получения одного отказа необходимо испытать выборку изделий из* 10 *приборов* ($x = 10$; $p = 0{,}05$). *Вычислить вероятность того, что для получения первого отказа понадобится испытать не более* 5 *приборов.*

Вероятность того, что для получения одного отказа необходимо испытать 10 приборов, равна $f(x;0{,}05) = 0{,}05 \cdot 0{,}95^9 = 0{,}031$.

Вероятность того, что для получения первого отказа понадобится испытать не более 5 приборов, равна $F(x;0{,}05) = 1-(1-0{,}05)^9 = 0{,}3697$.

1.2.6. Гипергеометрическое распределение

Описание, применение. Если в партии из N изделий находится D дефектных, то вероятность появления x дефектных изделий в выборке объема n изделий без возвращения будет подчинена гипергеометрическому распределению. Гипергеометрическое распределение широко применяется в задачах выборочного контроля качества продукции.

Свойства

Распределение вероятностей	$f(x;N,n,D) = \dfrac{c_D^x C_{N-D}^{n-x}}{c_N^n}$
Функция распределения	$F(x;N,n,D) = \sum\limits_{i=0}^{x} \dfrac{c_D^i C_{N-D}^{n-i}}{c_N^n}$
Среднее	$\mathbf{M}(x) = \dfrac{nD}{N}$
Дисперсия	$\mathbf{D}(x) = \dfrac{nD}{N}\left(1-\dfrac{D}{N}\right)\left(\dfrac{N-n}{N-1}\right)$

Коэффициент вариации $v = \left(\dfrac{N-D}{nD}\dfrac{N-n}{N-1}\right)^{\frac{1}{2}}$

Коэффициент асимметрии $\alpha_3 = (N-2D)[nD(N-D)]^{-\frac{1}{2}}\left(\dfrac{N-1}{N-n}\right)^{\frac{1}{2}}\dfrac{N-2n}{N-2}$

Коэффициент эксцесса

$$\alpha_4 = \frac{N^2(N-1)}{(N-2)(N-3)n(N-n)}\left\{\frac{N(N+1)-6N(N-n)}{D(N-D)} + \frac{3n(N-n)(N+6)}{N^2} - 6\right\}$$

Таблицы гипергеометрического распределения приведены в [23–25]. Наиболее полные таблицы опубликованы в [111, 112].

При расчетах полезны следующие рекуррентные формулы

$$f(x+1; N, D, n) = f(x; N, D, n)\left(\frac{n-x}{x+1}\right)\left(\frac{D-x}{N-n-D+x+1}\right);$$

$$f(x; N, D, n) = f(x; N, n, D) = f(D-x; N, D, N-n) = f(n-x; N, N-D, n) = \\ = f(N-n-D+x; N, N-D, N-n);$$

$$F(x; N, D, n) = F(x; N, n, D) = 1 - F(n-x-1; N, N-D, n) = \\ = 1 - F(D-x-1; N, D, N-n) = F(N-n-D+x; N, N-D, N-n).$$

Поскольку таблицы гипергеометрического распределения громоздки, а при больших значениях параметров N, n и D они отсутствуют, применяются различные аппроксимации этого распределения.

При $\dfrac{nD}{N} \to \infty$ и $\mathbf{D}(x) > 9$ [52]

$$F(x; N, n, D) = \Phi\left\{\frac{x + \frac{1}{2} - \mathbf{M}(x)}{\sqrt{\mathbf{D}(x)}}\right\},$$

где $\Phi(\ldots)$ — функция стандартного нормального распределения.

При $n < 0{,}1N$ и $D < 0{,}1N$ [51] гипергеометрическое распределение аппроксимируется распределением Пуассона (см. раздел 1.2.2) с параметром $\lambda = \dfrac{nD}{N}$. При $n < 0{,}1N$, $N \to \infty$ и фиксированном $\dfrac{D}{N}$ гипергеометрическое распределение аппроксимируется биномиальным (см. раздел 1.2.1) с параметрами n и $p = \dfrac{D}{N}$. Указанные аппроксимации действуют в различных диапазонах изменения параметров и не заменяют друг друга.

В [25] приведена аппроксимация с помощью бета-распределения (см. раздел 1.1.7), удовлетворительная при всех $N \geqslant 25$ (независимо от D и N),

$$F(x; N, n, D) = I_{1-x'}(n' - x + c, x - c + 1),$$

где $I_{1-x}(a, b)$ — функция бета-распределения;

$$x' = \frac{N(n+D-1) - 2nD}{N(N-2)}; \quad c = \frac{nD(D-1)(n-1)}{(N-1)[(N-D)(N-n) + nD - N]};$$

$$n' = \frac{(N-2)^2}{N-1}\,\frac{nD(N-n)(N-D)}{[(N-D)(N-n) + nD - N][N(n+D-1) - 2nD]}.$$

Укажем более точные приближения:

— биномиальным распределением с параметрами D, $p = \dfrac{2n - x}{2N - D + 1}$ или (для более точной аппроксимации)

$$p = \frac{2n - x}{2N - D + 1} - \frac{2D\left(x + \frac{1}{2} - \frac{nD}{N}\right)}{3(2N - D + 1)^2};$$

— распределением Пуассона с параметром $\lambda = \dfrac{(2D - x)(2n - x)}{2(2N - D - n + 1)}$ при $\dfrac{D}{N} \leqslant \dfrac{n}{N} \leqslant$ $\leqslant 0{,}1$ и $\lambda = \dfrac{nD}{N} + \dfrac{1}{3N}\left(\dfrac{nD}{N} - x\right)\left(2n - D + \dfrac{10Dn}{N}\right)$ при $0{,}1 < \dfrac{n}{N}$;

— нормальным распределением

$$F(x; N, n, D) = \begin{cases} \Phi\left(\dfrac{2}{\sqrt{N-1}}\left[\sqrt{(x+1)(N-D+x+1)} - \sqrt{(D-x)(n-x)}\right]\right) \\ \qquad \text{при} \quad 0{,}005 \leqslant p \leqslant 0{,}05 \quad \text{и} \quad 0{,}95 \leqslant p \leqslant 0{,}995; \\ \Phi\left\{\dfrac{2}{\sqrt{N}}\left[\sqrt{\left(x + \dfrac{3}{4}\right)\left(N - D - n + x + \dfrac{3}{4}\right)} - \right.\right. \\ \left.\left. - \sqrt{\left(D - x - \dfrac{1}{4}\right)\left(n - x - \dfrac{1}{4}\right)}\right]\right\} \quad \text{при} \quad 0{,}05 \leqslant p \leqslant 0{,}95. \end{cases}$$

Сравнительная точность различных аппроксимаций гипергеометрического распределения изучена в [113].

Задача 43. *Дано*: $N = 20$, $x = 3$, $D = 10$ *и* $n = 10$. *Вычислить значение* $F(x; N, n, D)$.

Точное значение:

$$F(3; 20, 10, 10) = \sum_{i=0}^{3} \frac{c_{10}^{i} c_{10}^{10-i}}{c_{20}^{10}} = \frac{1}{184756}(1 + 100 + 2025 + 14400) = 0{,}0894477.$$

Биномиальное приближение:

$$F(3; 20, 10, 10) = \sum_{i=0}^{3} C_n^i \left(\frac{D}{N}\right)^i \left(1 - \frac{D}{N}\right)^{n-i} = \sum_{i=0}^{3} C_{10}^i \cdot 0{,}5^i \cdot 0{,}5^{10-i} = 0{,}171875.$$

Ошибка велика, так как при $p = 0{,}5$ биномиальная аппроксимация наименее удачна.

Более точное биномиальное приближение:

$$F(3; 20, 10, 10) = \sum_{i=0}^{3} C_{10}^i \left(\frac{20 - 3}{40 - 10 + 1}\right)^i \left(1 - \frac{20 - 3}{40 - 10 + 1}\right)^{10-i} = 0{,}10387.$$

Ошибка еще велика, несмотря на уточнение.

Еще более точное приближение:

$$p = \frac{2 \cdot 10 - 3}{2 \cdot 20 - 10 + 1} + \frac{2 \cdot 10 \cdot \left(3 + \frac{1}{2} - 10 \cdot \frac{10}{20}\right)}{3 \cdot (2 \cdot 20 - 10 + 1)^2} = 0{,}53798117;$$

$$F(3; 20, 10, 10) = \sum_{i=0}^{3} C_{10}^i \cdot 0{,}53798117^i \cdot 0{,}462018827^{10-i} = 0{,}1166096.$$

Точность недостаточна, так как диапазон аппроксимации не подходит для биномиального распределения.

Пуассоновское приближение:

$$\lambda = \frac{10 \cdot 10}{20} = 5 \quad \text{и}$$

$$F(3; 20, 10, 10) = \sum_{i=0}^{3} e^{-5} \cdot \frac{5^i}{i!} = 6{,}737947 \cdot 10^{-3} \cdot \left(1 + 5 + \frac{25}{2} + \frac{125}{6}\right) = 0{,}265025.$$

Приближение очень грубо.

Так как $\frac{n}{N} = \frac{10}{20} > 0{,}1$, применим уточненную аппроксимацию:

$$\lambda = \frac{10 \cdot 10}{20} + \frac{1}{60} \cdot \left(\frac{10-10}{20} - 3\right) \cdot \left(20 - 10 + \frac{10 \cdot 10 \cdot 10}{20}\right) = 7;$$

$$F(3; 20, 10, 10) = \sum_{i=0}^{3} e^{-7} \cdot \frac{7^i}{i!} = e^{-7} \cdot \left(1 + 7 + \frac{49}{2} + \frac{343}{6}\right) = 0{,}0817654 \quad (\delta = 8{,}6\%).$$

Это уже достаточно близко к точному значению.

Нормальное приближение:

$$F(3; 20, 10, 10) = \Phi\left(\frac{3 + \frac{1}{2} - 5}{\sqrt{5 \cdot (1 - 0{,}5) \cdot \frac{10}{10}}}\right) = \Phi(-1{,}307) =$$

$$= 1 - 0{,}852 \cdot \exp\left[-\left(\frac{-1{,}307 - 1{,}5774}{2{,}0637}\right)^{2{,}34}\right] = 0{,}095 \quad (\delta = 6{,}7\%).$$

Более точное нормальное приближение:

$$F(3; 20, 10, 10) = \Phi\left\{\frac{2}{\sqrt{20}}\left[\sqrt{\left(3 + \frac{3}{4}\right)\left(20 - 10 + 10 + 3 + \frac{3}{4}\right)} - \right.\right.$$

$$\left.\left. - \sqrt{\left(10 - 3 - \frac{1}{4}\right)\left(10 - 3 - \frac{1}{4}\right)}\right]\right\} = \Phi(-1{,}3416) = 0{,}090 \quad (\delta = 1{,}1\%).$$

Бета-приближение:

$$x' = \frac{20(10 + 10 - 1) - 2 \cdot 10 \cdot 10}{20 \cdot 18} = 0{,}5; \quad c = \frac{10 \cdot 10 \cdot 9 \cdot 9}{19 \cdot (10 \cdot 10 + 100 - 20)} - 2{,}368;$$

$$n' = \frac{18^2}{19} \cdot \frac{10 \cdot 10 \cdot (20 - 10) \cdot (20 - 10)}{[(20 - 10) \cdot (20 - 10) = 10 \cdot 10 - 20] \cdot [20 \cdot (10 + 10 - 1) - 2 \cdot 10 \cdot 10]} = 5{,}263;$$

$$F(3; 20, 10, 10) = I_{0,5}(5{,}263 - 3 + 2{,}368; 3 - 2{,}368 + 1) = I_{0,5}(4{,}631; 1{,}632) = 0{,}0938$$
$$(\delta = 5{,}4\%).$$

ГЛАВА 2

ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ ВЕРОЯТНОСТЕЙ

Общие положения. В первой главе, посвященной анализу различных распределений вероятностей случайных величин, мы рассмотрели множество примеров применения математической статистики к решению практических задач. Однако в реальной жизни практически никогда не бывает так, чтобы исследователь располагал точным знанием закона распределения вероятностей наблюдаемых случайных величин. Ему в общем случае неизвестны как сам закон распределения вероятностей, так и его параметры. В распоряжении исследователя имеется лишь совокупность результатов наблюдений, и, основываясь только на них, он должен сделать выводы о параметрах распределения, если вид закона распределения вероятностей ему известен. Если же нет, то и сам закон распределения вероятностей ему придется выбирать на основании выборочных результатов наблюдений. В настоящей главе мы рассмотрим методы оценки параметров различных, заранее определенных по форме, распределений вероятностей случайных величин.

Различают два вида оценок параметров — *точечные* и *интервальные*. Предположим, оценке подлежит параметр γ некоторого распределения вероятностей по выборочным данным $x_1, x_2, \ldots, x_n$ некоторой случайной величины X. Точечной оценкой параметра γ по выборочным данным является некоторый функционал $\hat{\gamma}_n = \varphi(x_1, x_2, \ldots, x_n)$, позволяющий получить наилучшую оценку в принятых критериях.

В качестве критериев, характеризующих пригодность оценки параметра распределения, используются такие ее свойства, как *состоятельность*, *несмещенность*, *эффективность* и *достаточность*.

Оценка $\hat{\gamma}_n$ параметра γ является:

— состоятельной, если при $n \to \infty$ $\hat{\gamma}_n \to \gamma$;

— несмещенной, если $M(\hat{\gamma}_n) = \gamma$ (т. е. математическое ожидание оценки совпадает с ее истинным значением);

— достаточной, если оценка $\hat{\gamma}_n$ извлекает максимальную информацию из выборки;

— эффективной, если $D(\hat{\gamma}_n) = \min$ (т. е. когда дисперсия оценки минимальна).

Под интервальной оценкой параметра γ понимается интервал, границы которого $\hat{\gamma}_n^{\text{н}}(\alpha)$ и $\hat{\gamma}_n^{\text{в}}(\alpha)$ являются функционалами от выборочных значений случайной величины, и который с заданной вероятностью α содержит оцениваемый параметр: $P\{\hat{\gamma}_n^{\text{н}}(\alpha) < \gamma < \hat{\gamma}_n^{\text{в}}(\alpha)\} = \alpha$. Вероятность α называется *доверительной вероятностью*, а оценки $\hat{\gamma}_n^{\text{н}}(\alpha)$ и $\hat{\gamma}_n^{\text{в}}(\alpha)$ — соответственно *нижней* и *верхней доверительными границами*. Интервал $[\hat{\gamma}_n^{\text{н}}(\alpha), \hat{\gamma}_n^{\text{в}}(\alpha)]$ называется *доверительным интервалом*. Если длина доверительного интервала $l(\alpha) = \hat{\gamma}_n^{\text{в}}(\alpha) - \hat{\gamma}_n^{\text{н}}(\alpha) = \text{const}$, то для состоятельных и несмещенных оценок $\alpha \to 1$ при $n \to \infty$. При фиксированном объеме выборки n, α будет тем больше, чем больше l.

Различают два вида интервальных оценок: *одно-* и *двусторонние*. При *двусторонней* оценке задаются обе границы доверительного интервала, так что

$$P\{\hat{\lambda}_n^{\text{н}}(\alpha) < \gamma < \hat{\gamma}_n^{\text{в}}(\alpha)\} = \alpha \quad \text{и} \quad P\{\gamma < \hat{\gamma}_n^{\text{н}}(\alpha)\} = \alpha'; \quad P\{\gamma > \hat{\gamma}_n^{\text{в}}(\alpha)\} = \alpha'',$$

где $\alpha' + \alpha'' = 1 - \alpha$. Если $\alpha' = \alpha'' = \dfrac{1-\alpha}{2}$, то двусторонний доверительный интервал называется *симметричным*. Для него справедливы соотношения

$$P\{\gamma < \hat{\gamma}_n^{\text{в}}(\alpha)\} = \frac{1+\alpha}{2} \quad \text{и} \quad P\{\gamma > \hat{\gamma}_n^{\text{н}}(\alpha)\} = \frac{1+\alpha}{2}.$$

При односторонних доверительных интервалах границы интервалов задаются так, чтобы

$$P\{\gamma < \hat{\gamma}_n^{\text{в}}(\alpha)\} = \alpha \quad \text{или} \quad P\{\gamma > \hat{\gamma}_n^{\text{н}}(\alpha)\} = \alpha.$$

Величина $(1-\alpha)$ — дополнение доверительной вероятности до единицы — называется *уровнем значимости*. Этим термином обозначается вероятность появления события, которую исследователь связывает с *неслучайным* (*значимым*) *событием*. Очевидно, что двусторонний интервал для симметричных распределений аналогичен одностороннему при удвоенном уровне значимости.

Перед изложением конкретных методов оценки приведем ряд практических соображений. Наиболее существенной характеристикой оценки параметра распределения является ее *эффективность*. Именно эта характеристика обычно используется для сравнения методов оценки параметров распределения между собой. Как правило, эффективность оценки сравнивается с эффективностью оценки параметра распределения методом максимального правдоподобия (т. е. с наиболее эффективной оценкой). Легко видеть, что применение менее эффективных оценок (требующих, как правило, меньшего объема вычислений) может быть скомпенсировано соответствующим увеличением объема выборки.

И, наконец, поясним практический смысл процедуры оценки параметров распределения вероятностей. Так как само распределение наблюдаемых случайных величин является для исследователя той совокупностью данных, которой он располагает относительно наблюдаемого процесса, то и параметры распределения позволяют судить об основных чертах этого процесса. Например, когда мы спрашиваем, какова долговечность прибора, мы, по сути, ставим задачу оценки среднего значения (или математического ожидания) наблюдаемого распределения показателей долговечности. Если нас интересует, насколько стабилен наблюдаемый технологический процесс, то ответ на этот вопрос требует оценки разброса (рассеяния) наблюдаемых случайных величин, характеризующих качество технологического процесса.

2.1. Оценка параметров нормального распределения

Напомним, что плотность вероятностей нормально распределенной случайной величины описывается формулой

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

где μ и σ — параметры распределения, совпадающие со средним значением и среднеквадратическим отклонением.

2.1.1. Оценка среднего значения (μ)

2.1.1.1. Точечные оценки

2.1.1.1.1. Оценка максимального правдоподобия

Вычисляется по формуле $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Оценка — состоятельная, несмещенная, эффективная, достаточная и распределена как случайная величина также нормально со средним $M(\bar{x}) = \mu$ и дисперсией $D(\bar{x}) = \frac{\sigma^2}{n}$. Оценка максимального правдоподобия для случая выборок малого объема ($n \leqslant 10$) может быть модифицирована в форме [116]

$$\bar{x}^* = \sum_{i=1}^{n} \frac{x_i}{d_i \sum_{i=1}^{n} \frac{1}{d_i}}, \quad \text{где} \quad d_i = \sum_{j=1}^{n}(x_j - x_i)^2.$$

Эта форма позволяет несколько стабилизировать оценку в области центра группирования данных.

2.1.1.1.2. Оценка с помощью медианы

В качестве оценки μ может быть использована выборочная медиана

$$\tilde{x} = \begin{cases} \frac{1}{2}\left(x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n}{2}+1\right]}\right), & \text{если } \frac{n}{2} \text{ — целое;} \\ x_{\left[\frac{n+1}{2}\right]}, & \text{если } \frac{n+1}{2} \text{ — целое,} \end{cases}$$

где $x_{[k]}$ — k-я порядковая статистика, равная k-му по величине значению выборочного ряда $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$, ранжированного по возрастанию (будьте внимательны к тексту! — иногда квадратные скобки опускаются и x_k обозначает k-ю порядковую статистику).

Эффективность этой оценки при $n \to \infty$ равна $2/\pi = 0{,}637$, т. е. для того, чтобы эта оценка не уступала оценке максимального правдоподобия (см. раздел 2.1.1.1.1), необходим в $\pi/2 \approx 1{,}6$ раза больший объем выборки.

2.1.1.1.3. Оценки с помощью порядковых статистик

Поясним сначала смысл понятия „порядковая статистика". Как только любому члену наблюдаемого выборочного ряда ставится в соответствие его номер в упорядоченном по возрастанию ряду выборочных значений — этот член выборки становится порядковой статистикой. Для полного координирования порядковой статистики необходимо указать объем выборки и номер статистики. Впредь для того, чтобы отличить просто член выборки x от порядковой статистики, будем при-

менять для ее обозначения символ $x_{[i,n]}$ — т. е. i-я порядковая статистика в выборке объема n.

Теперь рассмотрим оценки среднего значения с помощью порядковых статистик. Предельным случаем такой оценки является оценка максимального правдоподобия (см. раздел 2.1.1.1.1), когда в оценке участвуют все члены выборки. Вторым предельным случаем является оценка с помощью медианы (см. раздел 2.1.1.1.2), т. е. с помощью только одной порядковой статистики. Естественно, что в этот диапазон эффективности возможных оценок (от 1 до 0,637) будут укладываться все остальные возможные оценки, использующие порядковые статистики.

Эти оценки будут уступать по эффективности оценке максимального правдоподобия. Однако в большинстве случаев соответствующим увеличением объема выборки, а также подбором порядковых статистик, используемых для оценки, и их весового вклада в общую оценку можно обеспечить достаточно высокую эффективность таких оценок. При этом сохраняются основные достоинства оценок по порядковым статистикам — простота, легкость вычислений, возможность получения оценок при отсутствии некоторых выборочных значений и, что особенно важно, устойчивость таких оценок к отклонению от нормальности распределения, от засорения выборки аномальными наблюдениями.

Среди широко применяемых порядковых статистик напомним $x_{[1]}, x_{[n]}$ — экстремальные значения; $\omega = x_{[n]} - x_{[1]}$ — выборочный размах; $\tilde{x}$ — медиана; выборочные квантили ($x_{[n\lambda]+1}$ — выборочная λ-квантиль).

Предварительно рассмотрим способы вычисления математических ожиданий порядковых статистик из стандартного нормального распределения.

Если $p = \dfrac{i-0{,}375}{n+0{,}250}$, то p-квантиль стандартного нормального распределения будет аппроксимировать математическое ожидание i-й порядковой статистики в выборке объема n из $N(0,1)$, т. е. $M(x_{[i]}) = u_{\frac{i-0{,}375}{n+0{,}250}}$. Поэтому необходимые аппроксимации могут быть получены из соответствующих аппроксимаций для квантилей стандартного нормального распределения.

Приведем некоторые полезные аппроксимации, полученные из аппроксимаций нормальных квантилей.

Из аппроксимации 5 раздела 1.1.1 [32]

$$F(x) = \frac{1}{2}\left\{\left[1 - \exp\left(-\frac{2x^2}{\pi}\right)\right]^{\frac{1}{2}} + 1\right\}$$

имеем

$$M(x_{[i]}) = \left\{-\frac{\pi}{2}\ln\left[1-(2p-1)^2\right]\right\}^{\frac{1}{2}} =$$

$$= \left\{-2{,}177586 + 3{,}14159\ln(n+0{,}25) - 1{,}570796[\ln(n+0{,}625-i) + \ln(i-0{,}375)]\right\}^{\frac{1}{2}}$$

$$\text{при}\quad \frac{n+1}{2} \leqslant i \leqslant 0{,}975n + 0{,}68175.$$

Погрешность аппроксимации не более 0,114.

Из аппроксимации 15 раздела 1.1.1 [39] получаем

$$u_p = 4{,}91\left[p^{0,14} - (1-p)^{0,14}\right],$$

из чего следует

$$M(x_{[i]}) = 4{,}91(n+0{,}25)^{0,14}\left[(i-0{,}375)^{0,14} + (n-i+0{,}625)^{0,14}\right].$$

Из аппроксимации 16 раздела 1.1.1 [36] имеем

$$M(x_{[i]}) = 2{,}0637\left(\ln\frac{n+0{,}250}{n-i+0{,}625} - 0{,}16\right)^{0{,}4274} - 1{,}5774.$$

Ошибка этой аппроксимации не более 0,0008 при $i \geqslant \dfrac{n+1}{2}$.

И, наконец, из аппроксимации 19 раздела 1.1.1 [40]

$$u_p = \frac{1}{2a}\left[-b - \sqrt{b^2-4ac}\right], \quad \text{где} \quad a=-0{,}416; \quad b=-0{,}717; \quad c=-\ln(2p)$$

получаем

$$M(x_{[i]}) = -0{,}86178 + 1{,}2019\left[2 + 1{,}664\ln\left(\frac{n-i+0{,}625}{n+0{,}250}\right)\right]^{\frac{1}{2}}.$$

Отметим, что во всех случаях при $i < \dfrac{n+1}{2}$ абсолютное значение $M(x_{[i]})$ сохраняется (изменяется только знак).

Всегда

$$M(x_{[i]}) = 0 \quad \text{при} \quad i = \frac{n+1}{2}\ (n\text{ — нечетное}) \quad \text{и} \quad i = \frac{n}{2}\ (n\text{ — четное});$$

$$M(x_{[i]}) > 0 \quad \text{при} \quad i > \frac{n+1}{2}\ \left(\text{или } \frac{n}{2}\right); \qquad M(x_{[i]}) < 0 \quad \text{при} \quad i < \frac{n+1}{2}\ \left(\text{или } \frac{n}{2}\right).$$

Рассмотрим теперь разные оценки среднего нормального распределения с помощью порядковых статистик.

2.1.1.1.3.1. Простые оценки Диксона [117]

Диксон предложил две простые оценки:

— среднее из двух наилучших наблюдений

$$\bar{x} = \frac{1}{2}(x_{[i]} + x_{[j]});$$

— среднее из всех наблюдений, кроме двух крайних

$$\bar{x} = \frac{1}{n-2}\sum_{i=2}^{n-1} x_{[i]}.$$

В табл. 6 приведены рекомендуемые номера оптимальных статистик (i и j) для $n = 1(1)20$.

Таблица 6

Оптимальные порядковые статистики Диксона [117]

n	2	3	4	5	6	7	8	9	10	11
i	1	1	2	2	2	2	3	3	3	3
j	2	3	3	4	5	6	6	7	8	9
n	12	13	14	15	16	17	18	19	20	∞
i	4	4	4	4	5	5	5	6	6	$0{,}27n$
j	9	10	11	12	12	13	14	14	15	$0{,}73n$

Относительная эффективность первой оценки стремится к 0,81 при $n \to \infty$, вторая практически не уступает оценке максимального правдоподобия ($> 0{,}99$).

2.1.1.1.3.2. Оценка Огавы [118, 119]

Вычисляется по формуле

$$\bar{x}_o = \sum_{i=1}^{m} x_{[n\lambda_i+1]},$$

где m — число наблюдений ($m < n$), по которым проводится оценка μ; $[n\lambda_i + 1]$ — целое число, ближайшее справа к $(n\lambda_i + 1)$ (определяет номера порядковых статистик, используемых для оценки); k_i, λ_i — числовые коэффициенты, табулированные в [119] (приведены в табл. 7).

Таблица 7

Значения коэффициентов оценки Огавы для различного числа оптимально расположенных порядковых статистик

(λ_i — верхняя строка; k_i — нижняя строка)

m	i									
	1	2	3	4	5	6	7	8	9	10
2	0,270	0,730								
	0,500	0,500								
3	0,163	0,500	0,837							
	0,297	0,407	0,297							
4	0,107	0,351	0,649	0,893						
	0,197	0,308	0,308	0,197						
5	0,074	0,255	0,500	0,745	0,926					
	0,133	0,233	0,269	0,233	0,133					
6	0,055	0,195	0,395	0,605	0,805	0,945				
	0,099	0,181	0,220	0,220	0,181	0,099				
7	0,040	0,147	0,308	0,500	0,692	0,853	0,960			
	0,071	0,140	0,186	0,203	0,186	0,140	0,071			
8	0,031	0,115	0,247	0,412	0,588	0,753	0,885	0,969		
	0,049	0,111	0,155	0,178	0,178	0,155	0,111	0,049		
9	0,024	0,092	0,202	0,343	0,500	0,657	0,798	0,908	0,976	
	0,044	0,091	0,130	0,155	0,163	0,155	0,130	0,091	0,044	
10	0,020	0,076	0,167	0,288	0,427	0,573	0,712	0,833	0,924	0,980
	0,036	0,075	0,109	0,133	0,147	0,147	0,133	0,109	0,175	0,136

Оценка Огавы используется для нахождения μ по выборке большого объема с помощью нескольких ($m < n$), оптимальным образом отобранных порядковых статистик. При $n \geqslant 5$ эффективность оценки не уступает оценке $\bar{x}$.

2.1.1.1.3.3. Оценка Пирсона–Тьюки

В [120] предложены упрощенные оценки, основанные на расстоянии между процентными точками частотных кривых распределений. Отметим, что $\bar{p}$-точка кривой равна порядковой статистике $x_{[n\lambda_i]}$, где $\lambda_i = p$.

Для μ Пирсоном и Тьюки [120], в частности, предложена оценка

$$\hat{\mu} = x_{[0,5n]} + 0{,}185\Delta, \quad \text{где} \quad \Delta = x_{[0,95n]} + x_{[0,05n]} - 2x_{[0,5n]}.$$

2.1.1.1.3.4. Быстрые оценки Кенуя [121]

Среднеквартильный размах

$$\tilde{\tilde{x}} = \frac{1}{2}\left(x_{[0,25n]} + x_{[0,75n]}\right).$$

Эффективность оценки $\approx 1{,}21/n$, предельная эффективность равна 0,64. Метод достаточно груб и может рекомендоваться как быстрый приближенный способ.

Быстрая оценка по трем квантилям

$$\tilde{\tilde{x}} = 0{,}2x_{\left[\frac{n}{16}\right]} + 0{,}6x_{\left[\frac{n}{2}\right]} + 0{,}2x_{\left[\frac{15n}{16}\right]}.$$

Эффективность оценки $\approx 83\%$, метод достаточно устойчив к отклонениям от нормальности.

Быстрая оценка по пяти квантилям

$$\tilde{\tilde{x}} = \frac{1}{6}\left[x_{\left[\frac{n}{16}\right]} + x_{\left[\frac{n}{4}\right]} + 2x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{3n}{4}\right]} + x_{\left[\frac{15n}{16}\right]}\right].$$

Эффективность оценки $\approx 0{,}93$, метод нечувствителен к отклонениям от нормальности.

2.1.1.1.3.5. Оптимальные комплексные оценки, использующие общий набор порядковых статистик

По набору порядковых статистик может быть проведена оценка и стандартного отклонения (σ). Однако для этого требуется иной, отличный от используемого при оценке μ, набор порядковых статистик. Представляется полезным рассмотреть вопрос об использовании общего набора порядковых статистик для совместной оценки μ и σ. При этом должен быть удовлетворен некий комплексный критерий оптимальности такого набора. Такие оценки рассмотрены в [122, 123].

Укажем оценку по двум порядковым статистикам (напомним, одинаковым при оценке и μ и σ), которая минимизирует линейную комбинацию дисперсий $D(\mu) + cD(\sigma)$ (см. табл. 8). Оценка имеет вид

$$\hat{\mu} = k(x_{[\alpha n]} + x_{[\beta n]}).$$

Аналогичная оценка по четырем порядковым статистикам имеет вид

$$\hat{\mu} = k_1\left(x_{[\alpha n]} + x_{[\beta n]}\right) + k_2\left(x_{[\alpha' n]} + x_{[\beta' n]}\right).$$

Она минимизирует сумму дисперсий $D(\mu) + D(\sigma)$. Необходимые константы оценок приведены в табл. 9.

Таблица 8

Оценка μ по двум статистикам [122]

c	k	a	b	Эффективность
1	0,5	0,1525	0,8475	0,729
2	0,5	0,1274	0,8726	0,683
3	0,5	0,1174	0,8853	0,594

Таблица 9

Оценка μ по четырем совместным квантилям [122]

Вариант оценки	k_1	k_2	α	α'	β	β'	Эффективность
1	0,1414	0,3586	0,0688	0,2912	0,9332	0,7088	0,9080
2	0,1097	0,4029	0,0389	0,2160	0,9611	0,7840	0,8570

2.1.1.1.3.6. Устойчивая (робастная) оценка Ходжеса–Лемана по средним Уолша

Определим в выборке $x_1, \ldots, x_n$ набор из $\frac{n(n+1)}{2}$ средних вида $z_{ij} = \frac{x_i + x_j}{2}$ $(i \leqslant j)$, называемых средними Уолша [124].

Оценка Ходжеса–Лемана [124] определяется как медиана средних Уолша, т. е. медиана ряда $z_1 \leqslant z_2 \leqslant \ldots \leqslant z_{\frac{n(n+1)}{2}}$. Следует отметить высокую устойчивость этой оценки к отклонениям от нормальности распределения и засоренности выборки аномальными наблюдениями.

2.1.1.1.4. Упрощенная оценка по шаблону [125]

Произвольно, но симметрично относительно предполагаемого среднего значения, выберем два числа a и b, определяющих размер шаблона $(a < b)$. Подсчитаем количество значений m, для которых $x \leqslant a$, и количество значений l, для которых $x \leqslant b$.

Для величин $p_1 = \frac{m}{n}$ и $p_2 = \frac{l}{n}$ находим по табл. 10 значение коэффициента $k(p_1, p_2)$. Искомая оценка определяется по формуле

$$\bar{x} = a - (b - a)k(p_1, p_2).$$

Задача 44. *В результате испытаний 30 приборов получены следующие значения ресурсной наработки:*

$$x_i = 721,\ 741,\ 752,\ 761,\ 763,\ 780,\ 794,\ 840,\ 890,\ 911,\ 944,\ 960,\ 961,\ 967,\ 1010,\ 1011,\\ 1012,\ 040,\ 1090,\ 1096,\ 1111,\ 1120,\ 1240,\ 1340,\ 1341,\ 1390,\ 1411,\ 1420,\ 1445,\ 1512.$$

Вычислить различными методами среднее значение ресурсной наработки.

Оценка максимального правдоподобия

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{30}\sum_{i=1}^{30} x_i = 1045{,}8.$$

Оценка с помощью медианы

$$\frac{n}{2} = 15;\ \frac{n+2}{2} = 16 \quad \text{и} \quad \tilde{x} = \frac{x_{[15]} + x_{[16]}}{2} = \frac{1010 + 1011}{2} = 1010{,}5.$$

Простые оценки Диксона (2.1.1.1.3.1)

Имеем $0{,}27n = 8{,}1$ и $0{,}73n = 21{,}9$. Следовательно, оценку будем проводить по 9-й и 22-й порядковым статистикам, т. е.

$$\bar{x} = \frac{x_{[9]} + x_{[22]}}{2} = \frac{890 + 1120}{2} = 1005 \quad \text{и} \quad \bar{x} = \frac{1}{n-2}\sum_{i=2}^{29} x_i = 1040{,}75.$$

Оценки Огавы (2.1.1.1.3.2)

— по двум статистикам:

$$[n\lambda_1 + 1] = [30 \cdot 0{,}27 + 1] = 10; \quad [n\lambda_2 + 1] = [30 \cdot 0{,}73 + 1] = 23;$$

$$\bar{x}_o = 0{,}5(x_{[10]} + x_{[23]}) = \frac{911 + 1240}{2} = 1075{,}5;$$

— по трем статистикам:

$$[n \cdot \lambda_1 + 1] = [30 \cdot 0{,}163 + 1] = 9; \quad [n \cdot \lambda_2 + 1] = [30 \cdot 0{,}50 + 1] = 16;$$
$$[n \cdot \lambda_3 + 1] = [30 \cdot 0{,}837 + 1] = 27;$$

$$\bar{x}_o = 0{,}297 \cdot (x_{[9]} + x_{[27]}) + 0{,}407 \cdot x_{[16]} = 0{,}297 \cdot (890 + 1411) + 0{,}407 \cdot 1011 = 1094{,}874;$$

— по десяти статистикам:

$$[n\cdot\lambda_1+1]=[30\cdot 0{,}020+1]=2;\qquad [n\cdot\lambda_6+1]=[30\cdot 0{,}573+1]=19;$$
$$[n\cdot\lambda_2+1]=[30\cdot 0{,}076+1]=4;\qquad [n\cdot\lambda_7+1]=[30\cdot 0{,}712+1]=23;$$
$$[n\cdot\lambda_3+1]=[30\cdot 0{,}167+1]=7;\qquad [n\cdot\lambda_8+1]=[30\cdot 0{,}833=1]=26;$$
$$[n\cdot\lambda_4+1]=[30\cdot 0{,}288+1]=10;\qquad [n\cdot\lambda_9+1]=[30\cdot 0{,}924+1]=29;$$
$$[n\cdot\lambda_5+1]=[30\cdot 0{,}427+1]=14;\qquad [n\cdot\lambda_{10}+1]=[30\cdot 0{,}98+1]=30.$$

Воспользовавшись коэффициентами из табл. 7, вычисляем

$$\bar{x}_o=0{,}036(x_2+x_{30})+0{,}075(x_4+x_{29})+0{,}109(x_7+x_{26})+0{,}133(x_{10}+x_{23})+$$
$$+0{,}147(x_{14}+x_{19})=0{,}036\cdot(741+1512)+0{,}075\cdot(761+1445)+0{,}109\cdot(794+1390)+$$
$$+0{,}133\cdot(911+1240)+0{,}147\cdot(967+1090)=1073{,}076.$$

Оценка Пирсона–Тьюки (2.1.1.1.3.3)

Имеем $[0{,}5n]=[0{,}5\cdot 30]=15$.
Далее $\Delta=x_{[28]}+x_{[2]}-2\cdot x_{[15]}=1420+741-2\cdot 1010=141$.
Окончательно имеем $\hat{\mu}=x_{[15]}+0{,}185\cdot 141=1010+0{,}185\cdot 141=1036{,}085$.

Быстрые оценки Кенуя (2.1.1.1.3.4)

— оценка по двум квантилям:

$$[0{,}25\cdot n]=[0{,}25\cdot 30]=8;\quad [0{,}75\cdot n]=[0{,}75\cdot 30]=23;$$
$$\tilde{\tilde{x}}=\frac{x_8+x_{23}}{2}=\frac{840+1240}{2}=1040;$$

— оценка по трем квантилям:

$$x_{\left[\frac{n}{16}\right]}=x_{\left[\frac{30}{16}\right]}=x_2;\quad x_{[0{,}5\cdot n]}=x_{15};\quad x_{\left[\frac{15\cdot n}{16}\right]}=x_{29};$$
$$\tilde{\tilde{x}}=0{,}2\cdot 741+0{,}6\cdot 1010+0{,}2\cdot 1445=1043{,}2;$$

— оценка по пяти квантилям:

$$\tilde{\tilde{x}}=\frac{1}{6}\cdot(x_2+x_8+2\cdot x_{15}+x_{23}+x_{29})=\frac{1}{6}\cdot(741+840+2\cdot 1010+1240+1445)=1047{,}67.$$

Оптимальные комплексные оценки (2.1.1.1.3.5)

— оценка по двум статистикам ($c=1$):

$$[0{,}1525\cdot n]=[0{,}1525\cdot 30]=5;\quad [0{,}8475\cdot n]=[0{,}8475\cdot 30]=26;$$
$$\hat{\mu}=0{,}5\cdot(x_5+x_{26})=0{,}5\cdot(763+1390)=1076{,}5;$$

— оценка по четырем статистикам ($c=1$):
(вариант 1)

$$[0{,}0668\cdot n]=[0{,}0668\cdot 30]=2;\quad [0{,}9332\cdot n]=[0{,}9332\cdot 30]=28;$$
$$[0{,}2912\cdot n]=[0{,}2912\cdot 30]=9;\quad [0{,}7088\cdot n]=[0{,}7088\cdot 30]=22;$$

$$\mu=0{,}1414\cdot(x_2+x_{28})+0{,}3586\cdot(x_9+x_{22})=$$
$$=0{,}1414\cdot(74+1420)+0{,}3586\cdot(890+1120)=1026{,}35;$$

(вариант 2)

$$[0{,}0389\cdot n]=[0{,}0389\cdot 30]=2;\quad [0{,}9611\cdot n]=[0{,}9611\cdot 30]=29;$$
$$[0{,}2160\cdot n]=[0{,}2160\cdot 30]=7;\quad [0{,}784\cdot n]=[0{,}784\cdot 30]=24;$$

$$\hat{\mu}=0{,}0971\cdot(x_2+x_{29})+0{,}4029\cdot(x_7+x_{24})=$$
$$=0{,}0971\cdot(741+1445)+0{,}4929\cdot(794+1340)=1072{,}049.$$

Оценка Ходжеса–Лемана (2.1.1.1.3.5)

Продемонстрируем вычисление оценки на примере пяти выборочных значений (полная оценка должна быть получена по $\frac{n(n+1)}{2} = \frac{30 \cdot 29}{2} = 435$ средним Уолша). Выберем для примера пять выборочных значений (вблизи среднего): x_i: 1012, 1040, 1090, 1096, 1111. Вычислим $\frac{n \cdot (n+1)}{2} = \frac{5 \cdot 4}{2} = 10$ средних Уолша:

$$z_\imath = 1026;\ 1051;\ 1054;\ 1061{,}5;\ 1065;\ 1068;\ 1075{,}5;\ 1093;\ 1100{,}5;\ 1103{,}5.$$

Далее вычисляем медиану значений $z_\imath$

$$\hat{\mu} = \operatorname{med}(z_\imath) = \frac{z_5 + z_6}{2} = \frac{1065 + 1068}{2} = 1066{,}5.$$

Упрощенная оценка по шаблону

Выбираем границы шаблона $a = 785$ и $b = 1115$. Имеем $m = 6$ и $l = 21$, откуда $p_1 = \frac{6}{30} = 0{,}2$ и $p_2 = \frac{21}{30} = 0{,}7$.

Из табл. 10 находим $k(0{,}2; 0{,}7) = -0{,}62$.

Следовательно, $\bar{x} = 785 - (1115 - 785) \cdot (-0{,}62) = 989{,}6$.

Для уточнения оценки попытаемся применить более узкий шаблон $a = 960{,}5$; $b = 1060$.

Тогда $m = 12$ ($P_1 = 0{,}4$); $l = 18$ ($P_2 = 0{,}6$) и $k(0{,}4; 0{,}6) = -0{,}50$.

Окончательно имеем $\bar{x} = 960{,}5 - (1060 - 960{,}5) \cdot (-0{,}50) = 1010{,}25$.

Таблица 10

Значения коэффициентов $k(p_1, p_2)$

p_1	p_2								
	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0,10	—	−5,33	−2,91	−2,10	−1,68	−1,44	−1,24	−1,11	−1,00
0,15	4,33	—	−5,20	−2,81	−2,00	−1,60	−1,32	−1,14	−1,00
0,20	1,91	4,20	—	−4,94	−2,62	−1,87	−1,42	−1,18	−1,00
0,25	1,10	1,81	3,94	—	−4,47	−2,39	−1,60	−1,24	−1,00
0,30	0,68	1,00	1,62	3,46	—	−4,00	−1,92	−1,33	−1,00
0,35	0,44	0,60	0,87	1,39	3,00	—	−2,79	−1,00	−1,00
0,40	0,24	0,32	0,42	0,60	0,93	1,79	—	−2,08	−1,00
0,45	0,11	0,14	0,18	0,24	0,33	0,50	1,08	—	−1,00
0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	—
0,55	−0,09	−0,11	−0,13	−0,16	−0,20	−0,25	−0,34	−0,50	−1,00
0,60	−0,16	−0,19	−0,23	−0,27	−0,32	−0,39	−0,50	−0,66	−1,00
0,65	−0,23	−0,27	−0,32	−0,37	−0,43	−0,50	−0,61	−0,75	−1,00
0,70	−0,28	−0,33	−0,38	−0,44	−0,50	−0,57	−0,68	−0,80	−1,00
0,75	−0,34	−0,39	−0,44	−0,50	−0,56	−0,63	−0,73	−0,84	−1,00
0,80	−0,40	−0,45	−0,50	−0,56	−0,62	−0,68	−0,77	−0,87	−1,00
0,85	−0,45	−0,50	−0,55	−0,61	−0,67	−0,73	−0,81	−0,89	−1,00
0,90	−0,50	−0,55	−0,60	−0,66	−0,71	−0,77	−0,84	−0,91	−1,00

Примечание. Для определения значений коэффициентов за пределами табличных значений p_1 и p_2 следует пользоваться соотношением $k(p_1, p_2) + k(p_2, p_1) = -1$.

2.1.1.2. Интервальные оценки

2.1.1.2.1. Оценка μ при известной дисперсии σ^2

Интервальная оценка с доверительной вероятностью α:

$$\mu_n^{\text{н}}(\alpha) = \bar{x} - u_\gamma \frac{\sigma}{\sqrt{n}}; \quad \mu_n^{\text{в}}(\alpha) = \bar{x} + u_\gamma \frac{\sigma}{\sqrt{n}},$$

где u_γ — γ-квантиль стандартного нормального распределения; $\gamma = \dfrac{1+\alpha}{2}$ для двусторонней оценки, $\gamma = \alpha$ для односторонней оценки.

2.1.1.2.2. Оценка μ при неизвестной дисперсии

$$\mu_n^{\text{н}}(\alpha) = \bar{x} - t_\gamma \frac{s}{\sqrt{n}}; \quad \mu_n^{\text{в}}(\alpha) = \bar{x} + t_\gamma \frac{s}{\sqrt{n}},$$

где t_γ — γ-квантиль распределения Стьюдента с $f = n - 1$ степенями свободы; $s^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$; $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$; $\gamma = \dfrac{1+\alpha}{2}$ для двусторонней оценки, $\gamma = \alpha$ для односторонней оценки.

Для квантилей стандартного нормального распределения и распределения Стьюдента рекомендуется использовать аппроксимации, приведенные в разделе 1.1.9:

$$u_\gamma = 4{,}91[\gamma^{0,14} - (1-\gamma)^{0,14}]; \quad u_\gamma = 2{,}0637\left(\ln\frac{1}{\gamma} - 0{,}16\right)^{0,4274} - 1{,}5774;$$

$$u_\gamma = \frac{0{,}717 - [0{,}717^2 - 40{,}8116\ln 2(1-\gamma)]^{\frac{1}{2}}}{-20{,}416} = \\ = -0{,}8617788 + 1{,}20192[0{,}846758 - 1{,}664\ln 2(1-\gamma)]^{\frac{1}{2}};$$

$$t_\gamma(f) = u_\gamma\left(1 - \frac{u_\gamma^2 + 1}{4f}\right)^{-1}; \quad t_\gamma(f) = \left\{f\exp\left(\frac{u_\gamma^2}{0{,}9975f - 0{,}445}\right) - 1\right\}^{\frac{1}{2}};$$

$$t_{0,975} = 1{,}96 + \frac{2{,}5}{f - 0{,}8}; \quad t_{0,975} = 2\sqrt{\frac{f}{f-2}}.$$

2.1.1.2.3. Оценка по выборочному размаху

$$\mu_n^{\text{н}}(\alpha) = \bar{x} - G_\gamma\omega; \quad \mu_n^{\text{в}}(\alpha) = \bar{x} + G_\gamma\omega,$$

где $\omega = x_{\max} - x_{\min}$ — выборочный размах; G_γ — γ-квантиль распределения размаха; $\gamma = \dfrac{1+\alpha}{2}$ для двусторонней и $\gamma = \alpha$ для односторонней оценки. Значения G_γ табулированы, например, в [24, 119]. Оценку рекомендуется применять при $n \leqslant 20$ (при дальнейшем росте объема выборки эффективность оценки резко падает).

Некоторые критические значения G_γ приведены в табл. 11.

Таблица 11

Критические значения G_γ [119]

n	γ				n	γ			
	0,95	0,975	0,99	0,995		0,95	0,975	0,99	0,995
2	3,157	6,353	15,910	31,000	12	0,158	0,194	0,241	0,277
3	0,885	1,304	2,111	8,800	13	0,147	0,181	0,224	0,256
4	0,529	0,717	1,023	1,316	14	0,138	0,170	0,209	0,239
5	0,388	0,507	0,685	0,843	15	0,131	0,160	0,197	0,224
6	0,312	0,399	0,523	0,628	16	0,124	0,151	0,186	0,212
7	0,263	0,333	0,429	0,507	17	0,118	0,144	0,177	0,201
8	0,230	0,288	0,366	0,429	18	0,113	0,137	0,168	0,191
9	0,205	0,255	0,322	0,374	19	0,108	0,131	0,161	0,182
10	0,186	0,230	0,288	0,333	20	0,104	0,126	0,154	0,175
11	0,170	0,210	0,263	0,302					

2.1.1.2.4. Оценка по интерквартильной широте

Дэвид и Джонсон [119] предложили интервальную оценку μ, основанную на выборочной медиане $\tilde{x}$ (см раздел 2.1.1.1.2) и интерквартильной широте l, равной разности между верхним и нижним выборочными квартилями, которые для выборок объема $n = 3 + 4k$ определяются порядковыми статистиками с номерами

$$x_{\frac{3(n+1)}{4}} \quad \text{и} \quad x_{\frac{n+1}{4}} \quad \left(l = x_{\frac{3(n+1)}{4}} - x_{\frac{n+1}{4}}\right).$$

Оценка имеет вид

$$\mu_n^{\text{н}}(\alpha) = \tilde{x} - R_\gamma \cdot l; \quad \mu_n^{\text{в}}(\alpha) = \tilde{x} + R_\gamma \cdot l,$$

где l — интерквартильная широта; R_γ — γ-квантиль распределения l; $\gamma = \dfrac{1+\alpha}{2}$ для двусторонней оценки, $\gamma = \alpha$ для односторонней оценки. Значения R_γ приведены в табл. 12.

Таблица 12

Значения R_γ [119]

n	γ			n	γ		
	0,95	0,975	0,99		0,95	0,975	0,99
11	0,470	0,623	0,876	35	0,260	0,319	0,393
15	0,400	0,514	0,676	39	0,246	0,301	0,369
19	0,354	0,448	0,573	43	0,234	0,286	0,349
23	0,321	0,402	0,506	47	0,224	0,273	0,332
27	0,296	0,367	0,458	51	0,215	0,261	0,317
31	0,276	0,341	0,422				

2.1.1.2.5. Оценка по среднему абсолютному отклонению

Если $C = \dfrac{1}{n}\sum\limits_{i=1}^{n}|x_i - \bar{x}|$ — среднее абсолютное отклонение, то интервальная оценка для среднего может быть определена, как

$$\mu_n^{\text{н}}(\alpha) = \bar{x} - K_\gamma C; \quad \mu_n^{\text{в}}(\alpha) = \bar{x} + K_\gamma C,$$

где K_γ – γ-квантиль распределения C (табулирована в [9, 127, 128]). Значения K_γ для $\gamma = 0{,}975$ приведены в табл. 13.

Таблица 13

Значения $K_{0,975}$

n	$K_{0,975}$	n	$K_{0,975}$	n	$K_{0,975}$
2	12,71	9	1,00	20	0,71
3	3,45	10	0,93	25	0,60
4	2,16	11	0,87	30	0,48
5	1,66	12	0,82	40	0,41
6	1,40	13	0,78	60	0,33
7	1,21	14	0,75	120	0,23
8	1,09				

2.1.1.2.6. Оценка 50%-го доверительного интервала по вероятному отклонению

$$\mu_n^{\text{н}}(0{,}5) = \bar{x} - 0{,}84535\frac{1}{n\sqrt{n-1}}\sum_{i=1}^{n}|x_i - \bar{x}|;$$

$$\mu_n^{\text{в}}(0{,}5) = \bar{x} + 0{,}84535\frac{1}{n\sqrt{n-1}}\sum_{i=1}^{n}|x_i - \bar{x}|.$$

Используется при $n \geqslant 7$ [129].

2.1.1.2.7. Интервальная оценка для медианы

При $n > 50$ доверительный интервал для медианы $\tilde{x}$ определяется порядковыми статистиками

$$x_k \leqslant \tilde{x} \leqslant x_{n-k+1},$$

где

$$k = \frac{1}{2}(n - 1{,}64\sqrt{n} - 1) \quad \text{при } \alpha = 0{,}9;$$

$$k = \frac{1}{2}(n - 1{,}96\sqrt{n} - 1) \quad \text{при } \alpha = 0{,}95;$$

$$k = \frac{1}{2}(n - 2{,}58\sqrt{n} - 1) \quad \text{при } \alpha = 0{,}99.$$

Для значений $n \leqslant 50$ номера порядковых статистик, заключающих в себе медиану, при $\alpha = 0{,}95$ и $\alpha = 0{,}99$ (двусторонние интервалы) приведены в табл. 14, заимствованной из [9].

Таблица 14

Доверительный интервал для медианы (номера порядковых статистик) [9]

n	α				n	α				n	α			
	0,95		0,99			0,95		0,99			0,95		0,99	
5	0	5	0	5	21	6	15	5	16	36	12	24	1	26
6	1	5	0	6	22	6	16	5	17	37	13	24	11	26
7	1	6	0	7	23	7	16	5	18	38	13	25	11	27
8	1	7	1	7	24	7	17	6	18	39	13	26	12	27
9	2	7	1	8	25	8	17	6	19	40	14	26	12	28
10	2	8	1	9	26	8	18	7	19	41	14	27	12	29
11	2	9	1	10	27	8	19	7	20	42	15	27	13	29
12	3	9	2	10	28	9	19	7	21	43	15	28	13	30
13	3	10	2	11	29	9	20	8	21	44	16	28	14	30
14	3	11	2	12	30	10	20	8	22	45	16	29	14	31
15	4	11	3	12	31	10	21	8	23	46	16	30	14	32
16	4	12	3	12	32	10	22	9	23	47	17	30	15	32
17	5	12	3	14	33	11	22	9	24	48	17	31	15	33
18	5	13	4	14	34	11	23	10	24	49	18	31	16	33
19	5	14	4	14	35	12	23	10	25	50	18	32	16	34
20	6	14	4	16										

При симметричном непрерывном распределении, каковым является нормальное, справедливы предложенные в [119] формулы односторонних доверительных интервалов для медианы:

— при доверительной вероятности $\alpha = 1 - \dfrac{2^{k-1}}{2n}$:

$$\tilde{x} < \frac{1}{2}(x_n + x_{n+1-k}); \quad \tilde{x} > \frac{1}{2}(x_1 + x_k);$$

— при доверительной вероятности $\alpha = 1 - \dfrac{k}{2n}$:

$$\tilde{x} < \max\left[x_{n-1}, \frac{x_n + x_{n+1-k}}{2}\right]; \quad \tilde{x} > \min\left[x_2, \frac{x_1 + x_k}{2}\right];$$

— при доверительной вероятности $\alpha = 1 - \dfrac{2(k-1)}{2n}$ $(k > 1)$:

$$\tilde{x} < \max\left[\frac{x_{n-1} + x_{n-2}}{2}; \frac{x_n + x_{n+1-k}}{2}\right]; \quad \tilde{x} > \min\left[\frac{x_2 + x_3}{2}; \frac{x_1 + x_k}{2}\right];$$

— при доверительной вероятности $\alpha = 1 - \dfrac{1 + \frac{1}{2}k(k-1) + \frac{1}{24}k(k-1)(k-2)(k-3)}{2n}$:

$$\tilde{x} < \max\left[x_{n-4}; \frac{x_n + x_{n+1-k}}{2}\right]; \quad \tilde{x} > \min\left[x_5; \frac{x_1 + x_k}{2}\right];$$

— при доверительной вероятности

$$\alpha = 1 - \frac{k\left[1 + \dfrac{(k-1)(k-2)}{6} + \dfrac{(k-1)(k-2)(k-3)(k-4)}{120}\right]}{2n}:$$

$$\tilde{x} < \max\left[x_{n-5}; \frac{x_n + x_{n+1-k}}{2}\right]; \quad \tilde{x} > \min\left[x_6; \frac{x_k + x_1}{2}\right];$$

— при доверительной вероятности $\alpha = 1 - \dfrac{2(k-1) + \frac{1}{6}(k-1)(k-2)(k-3)}{2n}$:

$$\tilde{x} < \max\left[\frac{x_{n-3} + x_{n-4}}{2}; \frac{x_n + x_{n+1-k}}{2}\right]; \quad \tilde{x} > \min\left[\frac{x_4 + x_5}{2}; \frac{x_1 + x_k}{2}\right];$$

— при доверительной вероятности $\alpha = 1 - \dfrac{1 + \frac{k(k-1)}{2}}{2n}$:

$$\tilde{x} < \max\left[x_3; \frac{x_1 + x_k}{2}\right]; \quad \tilde{x} > \min\left[x_{n-2}; \frac{x_n + x_{n+1-k}}{2}\right].$$

Оценки эффективны при $k \leqslant \dfrac{3n}{4}$ в случае нормального распределения (при $n < 20$).

Задача 45. *В результате наблюдений был получен следующий ряд данных:*

x_i: 2,4; 3,2; 4,1; 4,2; 4,8; 5,1; 6,2; 7,0; 9,0; 9,6 $(n = 10)$.

Найти двустороннюю интервальную оценку среднего при доверительной вероятности $\alpha = 0{,}95$.

Оценка при неизвестной дисперсии

Находим $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i = 5{,}56$ и $s = \left[\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})2\right]^{\frac{1}{2}} = 2{,}3796$. Далее $\gamma = \dfrac{1+\alpha}{2} = \dfrac{1 + 0{,}95}{2} = 0{,}975$; $f = n - 1 = 10 - 1 = 9$; $t_{0,975} \approx 2\sqrt{\dfrac{9}{7}} = 2{,}267$.

Окончательно имеем

$$\mu_n^{\text{н}}(\alpha) = \mu_{10}^{\text{н}}(0{,}95) = 5{,}56 - 2{,}267 \cdot \frac{2{,}3796}{\sqrt{10}} = 3{,}854;$$

$$\mu_n^{\text{в}}(\alpha) = \mu_{10}^{\text{в}}(0{,}95) = 5{,}56 + 2{,}267 \cdot \frac{2{,}3796}{\sqrt{10}} = 7{,}266.$$

Таким образом, с вероятностью 0,95 значение параметра (среднего значения) находится в интервале $3{,}854 \leqslant \mu \leqslant 7{,}266$.

Если бы интервалы были односторонними ($\gamma = \alpha = 0{,}95$), то следовало бы использовать $t_{0,95}$ вместо $t_{0,975}$ и, имея в виду, что $u_{0,95} = 1{,}645$,

$$t_{0,95} = u_{0,95}\left(1 - \frac{u_{0,95}^2 + 1}{4f}\right)^{-1} = 1{,}834.$$

Тогда получаем

$$\mu_n^{\text{н}}(0{,}95) = 5{,}56 - \frac{1{,}8342{,}3796}{\sqrt{10}} = 3{,}938; \quad \mu_n^{\text{в}}(0{,}95) = 5{,}56 + \frac{1{,}8342{,}3796}{\sqrt{10}} = 6{,}94.$$

С вероятностью 0,95 имеем односторонние интервалы $\mu \leqslant 6{,}94$ или $\mu \geqslant 3{,}938$.

Оценка по выборочному размаху (2.1.1.2.3)

Вычисляем размах выборки $\omega = x_{\max} - x_{\min} = 9{,}6 - 2{,}4 = 7{,}2$.

Для $\gamma = \dfrac{1+\alpha}{2} = 0{,}975$ и $n = 10$ находим из табл. 11 $G_{0,975}(10) = 0{,}23$. Тогда

$$\mu_n^{\text{н}}(\alpha) = \bar{x} - G_\gamma \cdot \omega = 5{,}56 - 0{,}23 \cdot 7{,}2 = 3{,}904;$$
$$\mu_n^{\text{в}}(\alpha) = \bar{x} + G_\gamma \cdot \omega = 5{,}56 + 0{,}23 \cdot 7{,}2 = 7{,}216.$$

Следовательно, с вероятностью 0,95 имеем $3{,}904 \leqslant \mu \leqslant 7{,}216$.

Оценка по интерквартильной широте (2.1.1.2.4)

Для удобства демонстрации добавим еще один член выборки x_{11}=10,0.

Имеем $\dfrac{3\cdot(n+1)}{4} = \dfrac{3\cdot 12}{4} = 9$; $\dfrac{n+1}{4} = \dfrac{11+1}{4} = 3$; $l = x_9 - x_3 = 9 - 4{,}1 = 4{,}9$;

Медиана равна $x_{\frac{n+1}{2}} = x_6 = 5{,}1$. Для $\gamma = 0{,}975$ ($\alpha = 0{,}95$) из табл. 12 находим $R_{0,975}(11) = 0{,}623$.

Окончательно получаем

$$\mu_n^{\text{н}}(\alpha) = 5{,}1 - 0{,}623 \cdot 4{,}9 = 2{,}047; \quad \mu_n^{\text{в}}(\alpha) = 5{,}1 + 0{,}623 \cdot 4{,}9 = 8{,}153;$$
$$2{,}047 \leqslant \mu \leqslant 8{,}153.$$

Оценка по среднему абсолютному отклонению (2.1.1.2.5)

Находим

$$C = \frac{1}{10} \cdot \sum_{i=1}^{10} |x_i - \bar{x}| =$$
$$= \frac{1}{10} \cdot (3{,}16 + 2{,}36 + 1{,}46 + 1{,}36 + 0{,}76 + 0{,}46 + 0{,}64 + 1{,}44 + 3{,}44 + 4{,}04) = 1{,}912.$$

Из табл. 13 для $n = 10$ имеем $K_{0,975}(10) = 0{,}93$ и далее

$$\mu_n^{\text{н}}(\alpha) = 5{,}56 - 0{,}93 \cdot 1{,}912 = 3{,}782; \quad \mu_n^{\text{в}}(\alpha) = 5{,}56 + 0{,}93 \cdot 1{,}912 = 7{,}338.$$

Следовательно, доверительный интервал равен $3{,}782 \leqslant \mu \leqslant 7{,}338$.

Оценка 50%-го доверительного интервала (2.1.1.2.6)

Имеем $\sum\limits_{i=1}^{n} |x_i - \bar{x}| = 19{,}12$ и

$$\mu_n^{\text{н}}(\alpha) = 5{,}56 - \frac{0{,}84535 \cdot 19{,}12}{10 \cdot \sqrt{9}} = 5{,}021; \quad \mu_n^{\text{в}}(\alpha) = 5{,}56 + \frac{0{,}84535 \cdot 19{,}12}{10 \cdot \sqrt{9}} = 6{,}0987.$$

Интервальная оценка для медианы (2.1.1.2.7)

Из табл. 14 для $n = 10$ ($\alpha = 0{,}95$) находим, что медиана лежит между 2-й и 8-й порядковыми статистиками, т. е. между $x_2 = 3{,}2$ и $x_8 = 7{,}0$. Следовательно, $3{,}2 \leqslant \tilde{x} \leqslant 7{,}0$.

Для приближенного подсчета имеем

$$k = \frac{n - 1{,}96 \cdot \sqrt{n} - 1}{2} = \frac{10 - 1{,}96 \cdot \sqrt{10} - 1}{2} = 1{,}4.$$

Следовательно, $[k] = 2$; $[n - k + 1] = 10$, т. е. необходимо использовать статистики x_2 и x_{10}, что приводит к результату $3{,}2 \leqslant \tilde{x} \leqslant 9{,}6$.

Приведем примеры вычисления односторонних приближенных оценок для медианы:

$k = 1 \quad \left(\alpha = 1 - \dfrac{1}{2n} = 0{,}95\right)$

$$\tilde{x} < x_{10} = 9{,}6; \quad \tilde{x} > x_1 = 2{,}4;$$

$k = 2 \quad \left(\alpha = 1 - \dfrac{2}{2n} = 0{,}9\right)$

$$\tilde{x} < \max\left[\frac{x_9 + x_8}{2}; \frac{x_9 + x_{10}}{2}\right] = \max\left[\frac{9 + 7}{2}; \frac{9{,}6 + 9}{2}\right] = 9{,}3;$$

$$\tilde{x} > \min\left[\frac{x_2 + x_3}{2}; \frac{x_1 + x_2}{2}\right] = \min\left[\frac{3{,}2 + 4{,}1}{2}; \frac{3{,}2 + 2{,}4}{2}\right] = 2{,}8.$$

2.1.2. Оценка дисперсии σ^2 и стандартного отклонения σ

2.1.2.1. Точечные оценки

2.1.2.1.1. Оценка максимального правдоподобия

Подсчитывается по формуле $s^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$, где $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$.

Для расчетов удобна формула $s^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^2 - \bar{x}^2$, из которой следует, что оценка максимального правдоподобия σ^2 является разностью между среднеарифметическим квадратом результатов наблюдений и квадратом среднего арифметического.

Оценка — состоятельная, достаточная, эффективная, но при малых n не являющаяся несмещенной. При $n < 30$ рекомендуется использовать несмещенную оценку

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right).$$

Если вместо оценки $\bar{x}$ используется параметр μ (когда он известен заранее), то во всех случаях используется оценка с n в знаменателе.

2.1.2.1.2. Оценка σ по выборочной дисперсии s^2

Оценка $s = \sqrt{s^2}$, где s^2 — оценка максимального правдоподобия для дисперсии (см. раздел 2.1.2.1.1), является смещенной. Несмещенной оценкой для σ является [25]

$$\tilde{s} = \frac{s}{\lambda}, \quad \text{где} \quad \lambda = \sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\dfrac{n-1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)}; \quad \Gamma(\ldots) \text{ — гамма-функция.}$$

При $n > 30$ $\lambda \approx 1 - \dfrac{2}{8n+1}$. Значения λ при $n \leqslant 45$ приведены в табл. 15.

Таблица 15

Значения λ [25]

n	λ	n	λ	n	λ	n	λ	n	λ
1	0,7979	6	0,9594	11	0,9776	16	0,9845	25	0,9900
2	0,8862	7	0,9650	12	0,9794	17	0,9540	30	0,9917
3	0,9213	8	0,9693	13	0,9810	18	0,9862	35	0,9929
4	0,9400	9	0,9727	14	0,9823	19	0,9869	40	0,9938
5	0,9515	10	0,9753	15	0,9835	20	0,9876	45	0,9945

2.1.2.1.3. Оценка σ по среднему абсолютному отклонению

Вычисляется по формуле

$$s^* = \frac{1}{nc}\sum_{i=1}^{n}|x_i - \bar{x}|, \quad \text{где} \quad c = \left[\frac{2}{\pi}\left(1 - \frac{1}{n}\right)\right]^{\frac{1}{2}}.$$

При $n \to \infty$ $c \approx \sqrt{\frac{2}{\pi}}\left(1 - \frac{1}{2n-1}\right) \approx 0{,}79788 \cdot \frac{2(n-1)}{2n-1}$. В силу невысокой эффективности оценка используется при $n \leqslant 20$.

2.1.2.1.4. Оценка σ по выборочному размаху

Подсчитывается по формуле $s_p = \frac{\omega}{d}$, где $\omega = x_{\max} - x_{\min}$ — размах выборки; d — коэффициент, зависящий от n (его значения табулированы в [25]), фрагмент этой таблицы содержится в табл. 16.

Таблица 16

Значения $1/d$

n	$1/d$	n	$1/d$	n	$1/d$	n	$1/d$
2	0,8862	7	0,3698	12	0,3069	17	0,2787
3	0,5908	8	0,3562	13	0,2998	18	0,2747
4	0,4857	9	0,3367	14	0,2935	19	0,2711
5	0,4299	10	0,3249	15	0,2880	20	0,2677
6	0,3946	11	0,3152	16	0,2831		

При $n < 20$ эффективность оценки s_p практически не отличается от эффективности оценки s (см. раздел 2.1.2.1.2), поэтому эта оценка предпочтительнее при малых выборках.

С помощью размаха можно быстро оценить верхнюю границу стандартного отклонения: $\sigma \leqslant \frac{\omega}{2}$ [130]. При $n \geqslant 200$ в качестве грубой оценки σ можно использовать оценку $s = \frac{\omega}{6}$.

2.1.2.1.5. Упрощенная оценка σ по шаблону

По аналогии с оценкой μ для быстрой, но грубой оценки σ можно использовать метод шаблонов [125]. При этом оценка имеет вид $s_ш = (d_2 - d_1)f(p_1,p_2)$, где $d_2 > d_1$ — размеры шаблона; p_1, p_2 — доля x, меньших d_1 и d_2 соответственно; $f(p_1, p_2)$ — коэффициент, зависящий от p_1 и p_2 (табулирован в табл. 17).

Таблица 17

Значения коэффициента $f(p_1, p_2)$

p_1	p_2							
	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45
0,10	—	4,17	2,27	1,64	1,32	1,12	0,97	0,87
0,15	−4,16	—	5,00	2,70	1,92	1,54	1,27	1,10
0,20	−2,27	−5,00	—	5,08	3,12	2,22	1,68	1,41
0,25	−1,64	−2,70	−5,88	—	6,67	3,57	2,38	1,85
0,30	−1,32	−1,92	−3,12	−5,67	—	7,70	3,70	2,56
0,35	−1,12	−1,54	−2,22	−3,57	−7,69	—	7,14	3,85
0,40	−0,97	−1,26	−1,69	−2,38	−3,70	−7,14	—	8,33
0,45	−0,87	−1,10	−1,41	−1,85	−2,56	−3,85	−8,33	—
0,50	−0,78	−0,96	−1,19	−1,49	−1,92	−2,56	−4,00	−7,69
0,55	−0,71	−0,85	−1,03	−1,25	−1,54	−1,92	−2,63	−3,85
0,60	−0,65	−0,78	−0,92	−1,09	−1,30	−1,56	−2,00	−2,63
0,65	−0,60	−0,70	−0,81	−0,94	−1,10	−1,28	−1,56	−1,92
0,70	−0,56	−0,64	−0,74	−0,84	−0,96	−1,10	−1,30	−1,54
0,75	−0,51	−0,58	−0,66	−0,75	−0,84	−0,94	−1,09	−1,25
0,80	−0,47	−0,53	−0,60	−0,66	−0,75	−0,81	−0,92	−1,03
0,85	−0,43	−0,48	−0,53	−0,58	−0,64	−0,70	−0,73	−0,85
0,90	−0,39	−0,43	−0,47	−0,51	−0,56	−0,60	−0,65	−0,71

p_1	p_2							
	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,90
0,10	0,78	0,71	0,65	0,60	0,56	0,51	0,47	0,39
0,15	0,96	0,85	0,78	0,70	0,64	0,58	0,53	0,43
0,20	1,19	1,03	0,92	0,81	0,74	0,66	0,60	0,47
0,25	1,49	1,25	1,09	0,94	0,84	0,75	0,66	0,51
0,30	1,92	1,53	1,30	1,10	0,96	0,84	0,74	0,56
0,35	2,56	1,92	1,56	1,28	1,10	0,94	0,81	0,60
0,40	4,00	2,63	2,00	1,56	1,30	1,09	0,92	0,65
0,45	7,69	3,85	2,63	1,92	1,54	1,25	1,03	0,71
0,50	—	7,69	4,00	2,56	1,92	1,49	1,19	0,78
0,55	−7,69	—	8,33	3,85	2,56	1,85	1,41	0,87
0,60	−4,00	−8,33	—	7,14	3,70	2,38	1,69	0,97
0,65	−2,56	−3,85	−7,14	—	7,69	3,57	2,22	1,12
0,70	−1,92	−2,56	−3,70	−7,69	—	6,67	3,12	1,32
0,75	−1,49	−1,85	−2,38	−3,57	−6,67	—	5,88	1,64
0,80	−1,19	−1,41	−1,69	−2,22	−3,12	−5,88	—	2,27
0,85	−0,96	−1,10	−1,27	−1,54	−1,92	−2,70	−5,00	4,17
0,90	−0,78	−0,87	−0,97	−1,12	−1,32	−1,32	−2,20	—

2.1.2.1.6. Оценка с помощью порядковых статистик

2.1.2.1.6.1. Оптимальная линейная оценка [24, 25, 118, 119]

Находится по формуле $\hat{\sigma} = \sum_{i=1}^{n} k_i x_i$, где x_i — i-я порядковая статистика; k_i — коэффициенты оценки (табулированы в [24, 118, 119]).

Значения коэффициентов k_i приведены в табл. 18. Учитывая, что $k_{\frac{n+1}{2}} = 0$, $k_i = -k_{n+1-i}$, в таблице приведены значения для членов выборки от 1 до $n/2$ (n — четное) или $\frac{n-1}{2}$ (n — нечетное).

Таблица 18

Значения $k_i \cdot (-10^4)$ для оптимальной линейной оценки σ [119]

n	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
2	8862									
3	5908	0								
4	4539	1102								
5	3724	1352	0							
6	3175	1386	0432							
7	2778	1352	0625	0						
8	2476	1294	0713	0230						
9	2237	1233	0751	0360	0					
10	2044	1172	0763	0436	0142					
11	1883	1115	0760	0481	0234	0				
12	1748	1061	0749	0506	0294	0087				
13	1632	1013	0735	0520	0335	0164	0			
14	1532	0968	0717	0526	0362	0212	0070			
15	1444	0927	0699	0526	0379	0247	0122	0		
16	1366	0899	0681	0524	0391	0272	0160	0530		
17	1297	0854	0663	0519	0398	0290	0189	0094	0	
18	1235	0822	0645	0512	0401	0302	0211	0125	0041	
19	1178	0792	0628	0525	0402	0312	0228	0150	0074	0
20	1128	0765	0611	0497	0402	0318	0241	0169	0101	0033

Эффективность оценки практически не уступает оценке максимального правдоподобия (см. раздел 2.1.2.1.2).

2.1.2.1.6.2. Оценка Огавы [118, 119]

Аналогична оптимальной линейной оценке, но вычисляется по $m < n$ порядковым статистикам, выбранным исходя из получения максимальной эффективности оценок в больших выборках без увеличения объема вычислений.

Вычисляется по формуле $\hat{\sigma} = \sum_{i=1}^{m} \beta_i x_{[n\lambda_i]+1}$, где m — число выборочных порядковых статистик, по которым производится оценка; λ_i — коэффициенты, определяющие номера порядковых статистик, участвующих в оценке; β_i — весовые коэффициенты оценки.

Коэффициенты λ_i и β_i табулированы в [118, 119] и приведены в табл. 19 ($[n\lambda_i]$ — целое число, ближайшее справа к $n\lambda_i$, и $[n\lambda_i] + 1$ — порядковый номер наблюдения).

Таблица 19

Значения λ_i (верхняя строка) **и β_i** (нижняя строка) **для оценки Огавы** [119]

m	i					
	1	2	3	4	5	6
2	0,069	0,931				
	−0,674	0,674				
4	0,023	0,127	0,873	0,977		
	−0,115	−0,237	0,237	0,115		
6	0,011	0,056	0,171	0,829	0,944	0,990
	−0,056	−0,126	−0,181	0,181	0,126	0,056

2.1.2.1.6.3. Линейная оценка Даутона

Даутон [131] предложил простую, но весьма эффективную оценку для σ в виде

$$\hat{\sigma} = \frac{1{,}77245}{n(n-1)} \sum_{i=1}^{n} (2i - n - 1)x_i.$$

При $n \leqslant 10$ эффективность оценки равна 0,94 по сравнению с оценкой максимального правдоподобия.

2.1.2.1.6.4. Оценка по сумме подразмахов (оценка Диксона) [117]

Вычисляется по формуле

$$\hat{\sigma} = k_n \sum_i \omega_i,$$

где $\omega_i = x_{n+1-i} - x_i$ — i-й подразмах; k_n — коэффициент оценки. Значения k_n и номера используемых в оценке подразмахов табулированы в [118, 119] и приведены в табл. 20 (очевидно, что $\omega_1 = \omega$ — обычный размах). Там же приведены коэффициенты k оценки $\hat{\sigma}$ по размаху ($\hat{\sigma} = k\omega$, см. табл. 16).

Таблица 20

Оценки $\hat{\sigma}$ по подразмахам

n	k (по размаху)	Эффективность оценки по размаху	Оценка по сумме подразмахов (оценка Диксона)	Эффективность оценки по подразмахам
2	0,886	1,000	$0{,}8862\omega$	1,000
3	0,591	0,992	$0{,}5908\omega$	0,992
4	0,486	0,975	$0{,}4857\omega$	0,975
5	0,430	0,955	$0{,}4299\omega$	0,955
6	0,395	0,933	$0{,}2916(\omega + \omega_2)$	0,957
7	0,370	0,911	$0{,}2370(\omega + \omega_2)$	0,967
8	0,351	0,890	$0{,}2197(\omega + \omega_2)$	0,970
9	0,337	0,869	$0{,}2068(\omega + \omega_2)$	0,968
10	0,325	0,850	$0{,}1968(\omega + \omega_2)$	0,964
11	0,315	0,831	$0{,}1608(\omega + \omega_2 + \omega_4)$	0,967
12	0,307	0,814	$0{,}1524(\omega + \omega_2 + \omega_4)$	0,972
13	0,300	0,797	$0{,}1456(\omega + \omega_2 + \omega_4)$	0,975
14	0,294	0,781	$0{,}1399(\omega + \omega_2 + \omega_4)$	0,977
15	0,288	0,766	$0{,}1352(\omega + \omega_2 + \omega_4)$	0,977
16	0,283	0,751	$0{,}1311(\omega + \omega_2 + \omega_4)$	0,975
17	0,279	0,738	$0{,}1050(\omega + \omega_2 + \omega_3 + \omega_5)$	0,978
18	0,275	0,725	$0{,}1020(\omega + \omega_2 + \omega_3 + \omega_5)$	0,978
19	0,271	0,712	$0{,}0993(\omega + \omega_2 + \omega_3 + \omega_5)$	0,979
20	0,268	0,700	$0{,}10446(\omega + \omega_2 + \omega_4 + \omega_6)$	0,980

2.1.2.1.6.5. Оценка Джини [122]

Используется как быстрая оценка для σ и подсчитывается по формуле

$$\hat{\sigma} = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} |x_i - x_j| = \sum_{i=1}^{\left[\frac{n}{2}\right]} \frac{n - 2i + 1}{n(n-1)}\, \omega_i,$$

где ω_i — i-й подразмах.

Оценка не уступает оценке Даутона (см. 2.1.2.1.6.3).

2.1.2.1.6.6. Оптимальные комплексные оценки, использующие общий набор порядковых статистик

Подробнее о сути оценки см. в разделе 2.1.1.1.3.5. Укажем оценку по двум порядковым статистикам (напомним, одинаковым при оценке μ и σ), минимизирующую линейную комбинацию дисперсий $D(\mu) + cD(\sigma)$ (табл. 21):

$$\hat{\sigma} = k(x_{[\alpha n]} - x_{[\beta n]}).$$

Аналогичные оценки по четырем порядковым статистикам имеют вид ($c = 1$)

$$\hat{\sigma} = 0{,}2581(x_{[0{,}9332n]} - x_{[0{,}0668n]}) + 0{,}2051(x_{[0{,}7088n]} - x_{[0{,}2912n]})$$

или

$$\hat{\sigma} = 0{,}1787(x_{[0{,}9611n]} - x_{[0{,}0389n]}) + 0{,}2353(x_{[0{,}7840n]} - x_{[0{,}2160n]}).$$

Эффективности оценок равны соответственно 0,735 и 0,792.

Таблица 21

Оценка σ по двум статистикам [117]

c	k	α	β	Эффективность
1	0,4875	0,8475	0,1525	0,552
2	0,4391	0,8726	0,1274	0,594
3	0,4160	0,8853	0,1174	0,614

Задача 46. *Имеется набор результатов измерений*

$$x_i\colon 1{,}4;\ 2{,}1;\ 2{,}9;\ 3{,}1;\ 3{,}8;\ 4{,}1;\ 4{,}3;\ 4{,}6;\ 5{,}1;\ 6{,}1.$$

Необходимо вычислить оценку дисперсии и среднеквадратического отклонения различными способами.

Оценка максимального правдоподобия (2.1.2.1.1)

$$s^2 = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\cdot\sum_{i=1}^{n}x_i^2 - \left(\frac{1}{n}\cdot\sum_{i=1}^{n}x_i\right)^2 = \frac{1}{10}\cdot 158{,}51 - (3{,}75)^2 = 1{,}7885$$

или $s^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 = 1{,}9872$ (с использованием поправки на смещение).

Следует помнить, что при малых n смещение оценки может быть значительным.

Оценка по среднему абсолютному отклонению (2.1.2.1.3)

$$s* = \frac{1}{n\cdot c}\sum_{i=1}^{n}|x_i - \bar{x}|, \quad \text{где} \quad c = \sqrt{\frac{2}{\pi}\cdot\left(1 - \frac{1}{10}\right)} = 0{,}75694.$$

Имеем $\sum\limits_{i=1}^{n}|x_i - \bar{x}| = 11$ и $s^* = \dfrac{1}{10\cdot 0{,}75694}\cdot 11 = 1{,}4181$.

Оценка по выборочному размаху (2.1.2.1.4)

Имеем $\omega = x_{\max} - x_{\min} = 6{,}1 - 1{,}4 = 4{,}7$. Для $n = 10$ из табл. 16 получаем $\dfrac{1}{d} = 0{,}3249$ и $s_p = 4{,}7\cdot 0{,}3249 = 1{,}529$.

Упрощенная оценка по шаблону (2.1.2.1.5)

Выбираем $d_2 = 4{,}5$ и $d_1 = 2{,}5$, тогда $p_1 = 0{,}2$ и $p_2 = 0{,}8$. Из табл. 17 имеем $f(0{,}2; 0{,}8) = 0{,}60$ и $s_{ш} = (4{,}5 - 2{,}5)\cdot 0{,}6 = 1{,}20$.

Оптимальная линейная оценка (2.1.2.1.6.1)

Находим с помощью табл. 18

$$\hat{\sigma} = \sum_{i=1}^{n} k_i \cdot x_{[i]} = 0{,}2044 \cdot (6{,}1 - 1{,}4) + 0{,}1172 \cdot (5{,}1 - 2{,}1) + 0{,}0763 \cdot (4{,}6 - 2{,}9) + \\ + 0{,}0436 \cdot (4{,}3 - 3{,}1) + 0{,}0142 \cdot (4{,}1 - 3{,}8) = 1{,}49857.$$

Оценка Огавы (2.1.2.1.6.2)

Выполним оценку по четырем порядковым статистикам. В соответствии с табл. 19 их номера будут

$$[n \cdot \lambda_1] + 1 = [10 \cdot 0{,}023] + 1 = 1; \quad [n \cdot \lambda_2] + 1 = [10 \cdot 0{,}127] + 1 = 2;$$
$$[n \cdot \lambda_3] + 1 = [10 \cdot 0{,}873] + 1 = 9; \quad [n \cdot \lambda_4] + 1 = [10 \cdot 0{,}977] + 1 = 10.$$

Тогда имеем, используя коэффициенты из табл. 19,

$$\hat{\sigma} = -0{,}115 \cdot x_1 - 0{,}237 \cdot x_2 + 0{,}237 \cdot x_9 + 0{,}115 \cdot x_{10} = \\ = 0{,}115 \cdot (6{,}1 - 1{,}4) + 0{,}237 \cdot (5{,}1 - 2{,}1) = 1{,}2515.$$

Линейная оценка Даутона (2.1.2.1.6.3)

$$\hat{\sigma} = \frac{1{,}77245}{10 \cdot (10-1)} \cdot \sum_{i=1}^{n} (2i - n - 1) \cdot x_i = 0{,}019694 \cdot (-9 \cdot 1{,}4 - 7 \cdot 2{,}1 - 5 \cdot 2{,}9 - 3 \cdot 3{,}1 - \\ - 1 \cdot 3{,}8 + 1 \cdot 4{,}1 + 3 \cdot 4{,}3 + 5 \cdot 4{,}6 + 7 \cdot 5{,}1 + 9 \cdot 6{,}1) = 1{,}49083.$$

Оценка по сумме подразмахов (2.1.2.1.6.4)

Из табл. 20 (для $n = 10$) следует, что

$$\hat{\sigma} = 0{,}1968 \cdot (\omega + \omega_2) = 0{,}1968 \cdot (x_{10} - x_1 + x_9 - x_2) = \\ = 0{,}1968 \cdot (6{,}1 - 1{,}4 + 5{,}1 - 2{,}1) = 1{,}51536.$$

Оценка Джини (2.1.2.1.6.5)

Имеем

$$\sum_{i,j=1}^{n} |x_i - x_j| = \sum_{i=1}^{\left[\frac{n}{2}\right]} (n - 2i + 1) \cdot \omega_i = (n-1) \cdot \omega_1 + (n-3) \cdot \omega_2 + (n-5) \cdot \omega_3 + \\ + (n-7) \cdot \omega_4 + (n-9) \cdot \omega_5 = 75{,}7 \quad \text{и} \quad \hat{\sigma} = \frac{1}{10 \cdot 9} \cdot 75{,}7 = 0{,}842.$$

Из результата видно, что оценка очень груба.

Оценки, использующие общий набор порядковых статистик (2.1.2.1.6.6)

Имеем для двух статистик (пусть $c = 1$)

$$\hat{\sigma} = 0{,}4875 \cdot (x_{[0{,}8475 \cdot 10]} - x_{[0{,}1525 \cdot 10]}) = 0{,}4875 \cdot (x_8 - x_1) = 0{,}4875 \cdot (4{,}6 - 1{,}4) = 1{,}56;$$

оценка по четырем статистикам:

вариант 1

$$\hat{\sigma} = 0{,}2581 \cdot (x_{[0{,}9332 \cdot 10]} - x_{[0{,}0688 \cdot 10]}) + 0{,}2051 \cdot (x_{[0{,}7088 \cdot 10]} - x_{[0{,}2912 \cdot 10]}) = \\ = 0{,}2581 \cdot (x_9 - x_1) + 0{,}2051 \cdot (x_7 - x_3) = 0{,}2581 \cdot (5{,}1 - 1{,}4) + 0{,}2051 \cdot (4{,}3 - 2{,}9) = 1{,}2421.$$

вариант 2

$$\hat{\sigma} = 0{,}1787 \cdot (x_{[0{,}9611 \cdot 10]} - x_{[0{,}0389 \cdot 10]}) + 0{,}2353 \cdot (x_{[0{,}7840 \cdot 10]} - x_{[0{,}2160 \cdot 10]}) = \\ = 0{,}1787 \cdot (x_{10} - x_1) + 0{,}2353 \cdot (x_8 - x_2) = 0{,}1787 \cdot (6{,}1 - 1{,}4) + 0{,}2353 \cdot (4{,}6 - 2{,}1) = 1{,}428.$$

2.1.2.2. Интервальные оценки

2.1.2.2.1. Интервальные оценки дисперсии σ^2

Интервальные оценки при доверительной вероятности α равны

$$(s_n^2)_\alpha^{\text{н}} = \frac{1}{\chi^2_{\gamma'}} \sum_{i=1}^{n} (x_i - \bar{x})^2; \quad (s_n^2)_\alpha^{\text{в}} = \frac{1}{\chi^2_{\gamma''}} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

где χ^2_γ — γ-квантиль распределения χ^2 с $f = n-1$ степенями свободы (если параметр $\bar{x} \approx \mu$ известен, то $f = n$); $\gamma' = \frac{1+\alpha}{2}$ для двусторонней оценки и $\gamma' = \alpha$ для односторонней оценки; $\gamma'' = \frac{1-\alpha}{2}$ для двусторонней оценки и $\gamma'' = 1-\alpha$ для односторонней оценки.

Для аппроксимации χ^2_γ рекомендуется использовать формулу Вилсона–Хилферти [61]

$$\chi^2_\gamma(f) = f\left[1 - \frac{2}{9f} + u_\gamma\sqrt{\frac{2}{9f}}\right]^3,$$

где u_γ — γ-квантиль стандартного нормального распределения.

Для практически применяемых уровней достоверности $\alpha = 0{,}9$; 0,95 и 0,99 значения u_γ приведены в табл. 22.

Таблица 22

Значения u_γ

α	Односторонние границы				Двусторонние границы			
	γ'	γ''	$u_{\gamma'}$	$u_{\gamma''}$	γ'	γ''	$u_{\gamma'}$	$u_{\gamma''}$
0,90	0,90	0,10	1,28255	−1,28255	0,950	0,050	1,64485	−1,64485
0,95	0,95	0,05	1,64485	−1,64485	0,975	0,025	1,95996	−1,95996
0,99	0,99	0,01	2,32635	−2,32635	0,995	0,005	2,57582	−2,57582

2.1.2.2.2. Интервальная оценка σ по размаху

Оценка находится по формулам

$$s_{\text{н}}^{\text{н}}(\alpha) = \frac{x_{vax} - x_{\min}}{\omega(\gamma')}; \quad s_n^{\text{в}}(\alpha) = \frac{x_{\max} - x_{\min}}{\omega(\gamma'')},$$

где $\omega(\gamma)$ — γ-квантиль распределения размаха выборки объема n из стандартного нормального распределения (табулированы в [25, 29]); $\gamma' = \frac{1+\alpha}{2}$ для двусторонней и $\gamma' = \alpha$ для односторонней оценки; $\gamma'' = \frac{1-\alpha}{2}$ для двусторонней, $\gamma'' = 1-\alpha$ для односторонней оценки. Оценка применяется при $n \leqslant 20$. Для обычно применяемых величин $\alpha = 0{,}90$; 0,95 и 0,99 необходимые значения $\omega(\gamma)$ для $n = 1(1)20$ приведены в табл. 23.

2.1.2.2.3. Оценка по среднему абсолютному отклонению

Вычисляется по формулам

$$s_n^{\text{н}} = \frac{1}{nm(\gamma')} \sum_{i=1}^{n} |x_i - \bar{x}|; \quad s_n^{\text{в}} = \frac{1}{nm(\gamma'')} \sum_{i=1}^{n} |x_i - \bar{x}|,$$

Таблица 23

Значения квантилей распределения размаха $\omega(\gamma)$ [25]

n	Односторонние оценки						Двусторонние оценки					
	$\alpha = 0{,}90$		$\alpha = 0{,}95$		$\alpha = 0{,}99$		$\alpha = 0{,}90$		$\alpha = 0{,}95$		$\alpha = 0{,}99$	
	γ'	γ''	γ'	γ''	γ'	γ''	γ'	γ''	γ'	γ''	γ'	γ''
	0,90	0,10	0,95	0,05	0,99	0,01	0,95	0,05	0,975	0,025	0,995	0,005
3	2,90	0,62	3,31	0,43	4,12	0,19	3,31	0,43	3,68	0,30	4,42	0,13
4	3,24	0,98	3,63	0,76	4,40	0,43	3,63	0,76	3,98	0,59	4,69	0,34
5	3,48	1,26	3,86	1,03	4,60	0,66	3,86	1,03	4,20	0,85	4,89	0,55
6	3,66	1,49	4,03	1,25	4,76	0,87	4,03	1,25	4,36	1,06	5,03	0,75
7	3,81	1,68	4,17	1,44	4,88	1,05	4,17	1,44	4,49	1,25	5,15	0,92
8	3,93	1,83	4,29	1,60	4,99	1,20	4,29	1,60	4,61	1,41	5,26	1,08
9	4,04	1,97	4,39	1,74	5,08	1,34	4,39	1,74	4,70	1,55	5,34	1,21
10	4,13	2,09	4,47	1,86	5,16	1,47	4,47	1,86	4,79	1,67	5,42	1,33
11	4,21	2,20	4,55	1,97	5,23	1,58	4,55	1,97	4,86	1,78	5,49	1,45
12	4,29	2,30	4,62	2,07	5,29	1,68	4,62	2,07	4,92	1,88	5,54	1,55
13	4,35	2,39	4,68	2,16	5,35	1,77	4,68	2,16	4,99	1,97	5,60	1,64
14	4,41	2,47	4,74	2,24	5,40	1,86	4,74	2,24	5,04	2,06	5,65	1,72
15	4,47	2,54	4,80	2,32	5,45	1,93	4,80	2,32	5,09	2,14	5,70	1,80
16	4,52	2,61	4,85	2,39	5,49	2,01	4,85	2,39	5,14	2,21	5,74	1,88
17	4,57	2,67	4,89	2,45	5,54	2,07	4,89	2,45	5,18	2,27	5,78	1,94
18	4,61	2,73	4,93	2,51	5,57	2,14	4,93	2,51	5,22	2,34	5,82	2,01
19	4,65	2,79	4,97	2,57	5,61	2,20	4,97	2,57	5,26	2,39	5,85	2,07
20	4,69	2,84	5,01	2,62	5,65	2,25	5,01	2,62	5,30	2,45	5,89	2,12

Таблица 24

Значения квантилей распределения среднего абсолютного отклонения $m(\gamma)$ [25]

n	Односторонние оценки						Двусторонние оценки					
	$\alpha = 0{,}90$		$\alpha = 0{,}95$		$\alpha = 0{,}99$		$\alpha = 0{,}90$		$\alpha = 0{,}95$		$\alpha = 0{,}99$	
	γ'	γ''	γ'	γ''	γ'	γ''	γ'	γ''	γ'	γ''	γ'	γ''
	0,90	0,10	0,95	0,05	0,99	0,01	0,95	0,05	0,975	0,025	0,995	0,005
2	1,163	0,088	1,386	0,044	1,821	0,009	1,386	0,044	1,585	0,022	1,985	0,004
3	1,117	0,238	1,276	0,166	1,586	0,073	1,276	0,166	1,417	0,116	1,703	0,073
4	1,089	0,328	1,224	0,254	1,489	0,145	1,224	0,254	1,344	0,199	1,59	0,145
5	1,069	0,386	1,187	0,315	1,419	0,203	1,187	0,305	1,292	0,260	1,507	0,203
6	1,052	0,428	1,158	0,360	1,366	0,250	1,158	0,360	1,253	0,306	1,445	0,25
7	1,038	0,459	1,135	0,394	1,325	0,287	1,135	0,394	1,222	0,342	1,397	0,287
8	1,026	0,484	1,116	0,422	1,292	0,318	1,116	0,422	1,196	0,372	1,358	0,318
9	1,016	0,504	1,100	0,445	1,264	0,344	1,100	0,445	1,175	0,396	1,326	0,344
10	1,007	0,521	1,086	0,464	1,240	0,366	1,086	0,464	1,156	0,417	1,299	0,366

где $m(\gamma)$ — γ-квантиль распределения среднего абсолютного отклонения; $\gamma' = (1+\alpha)/2$, $\gamma'' = (1+\alpha)/2$ для двусторонней оценки и $\gamma' = \alpha$, $\gamma'' = 1 - \alpha$ для односторонней оценки.

Для $n = 1(1)100$ и $\alpha = 0{,}90$; $0{,}95$ и $0{,}99$ значения $m(\gamma)$ приведены в табл. 24.

2.1.2.2.4. Интервальная оценка σ, основанная на ее точечной оценке s

Оценка вычисляется по формулам

$$s_n^{\text{н}} = \sqrt{\frac{n-1}{\chi^2_{\gamma'}}}\,s; \quad s_n^{\text{в}} = \sqrt{\frac{n-1}{\chi^2_{\gamma''}}}\,s,$$

где $s = \left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{\frac{1}{2}}$; χ^2_γ — γ-квантиль распределения χ^2 с $f = n-1$ степенями свободы; $\gamma' = \frac{1+\alpha}{2}$ $(\gamma' = \alpha)$ для двусторонней (односторонней) оценки; $\gamma'' = \frac{1-\alpha}{2}$ $(\gamma'' = 1-\alpha)$ для двусторонней (односторонней) оценки. Значения $\sqrt{\frac{n-1}{\chi^2_\gamma}}$ табулированы в [25, 46] и приведены в табл. 25.

В [36] для $\alpha = 0{,}95$ приведены достаточно простые и точные эмпирические формулы:

Таблица 25

Значения $\sqrt{\frac{n-1}{\chi^2_\gamma}}$ для интервальной (двусторонней) оценки σ

$n-1$	$\alpha = 0{,}90$		$\alpha = 0{,}95$		$\alpha = 0{,}99$	
	$\gamma' = 0{,}95$	$\gamma'' = 0{,}05$	$\gamma' = 0{,}975$	$\gamma'' = 0{,}025$	$\gamma' = 0{,}995$	$\gamma'' = 0{,}005$
2	0,578	4,42	0,521	6,28	0,434	14,12
3	0,620	2,92	0,566	3,73	0,483	6,47
4	0,649	2,37	0,599	2,87	0,519	4,40
5	0,672	2,09	0,624	2,45	0,546	3,48
6	0,690	1,92	0,644	2,20	0,569	2,98
7	0,705	1,80	0,661	2,04	0,588	2,66
8	0,718	1,71	0,675	1,92	0,604	2,44
9	0,729	1,65	0,688	1,83	0,618	2,28
10	0,739	1,59	0,699	1,75	0,630	2,15
11	0,748	1,55	0,708	1,70	0,641	2,06
12	0,755	1,52	0,717	1,65	0,651	1,98
13	0,762	1,49	0,725	1,61	0,660	1,91
14	0,769	1,46	0,732	1,58	0,669	1,85
15	0,775	1,44	0,739	1,55	0,676	1,81
16	0,780	1,42	0,745	1,52	0,683	1,76
17	0,785	1,40	0,750	1,50	0,690	1,73
18	0,790	1,39	0,756	1,48	0,696	1,70
19	0,794	1,37	0,760	1,46	0,702	1,67
20	0,798	1,36	0,765	1,44	0,707	1,64
22	0,805	1,34	0,773	1,42	0,717	1,60
24	0,812	1,32	0,781	1,39	0,726	1,56
26	0,818	1,30	0,788	1,37	0,734	1,53
28	0,823	1,29	0,794	1,35	0,741	1,50
30	0,828	1,27	0,799	1,34	0,748	1,48
35	0,838	1,25	0,811	1,30	0,762	1,43
40	0,847	1,23	0,821	1,28	0,774	1,39
45	0,854	1,21	0,829	1,26	0,784	1,36
50	0,861	1,20	0,837	1,24	0,793	1,34
55	0,866	1,19	0,843	1,23	0,801	1,32
60	0,871	1,18	0,849	1,22	0,808	1,30
70	0,879	1,16	0,858	1,20	0,820	1,27
80	0,886	1,15	0,866	1,18	0,829	1,25
90	0,892	1,14	0,873	1,17	0,838	1,23
100	0,897	1,13	0,879	1,16	0,845	1,22

— для нижней границы

$$\sqrt{\frac{n-1}{\chi^2_{0,975}}} = \frac{\sqrt{2n-1{,}74}}{1 < 96 + \sqrt{2n-2}};$$

— для верхней границы

$$\sqrt{\frac{n-1}{\chi^2_{0,025}}} = \begin{cases} 11{,}54(n-3{,}61)^2 + 1{,}98; & 2 \leqslant n \leqslant 4; \\ \dfrac{\sqrt{2n-0{,}47}}{\sqrt{2n-0{,}8}-1{,}96}; & n \geqslant 5. \end{cases}$$

Оценки улучшаются, если вместо s использовать несмещенную оценку $s^* = ks$, где $k = 1 + \dfrac{0{,}254}{n-1}$ [36].

Отметим, что в нашем случае рассматриваются симметричные квантили $\gamma' = 0{,}95$ и $\gamma'' = 0{,}05$, $\gamma' = 0{,}975$ и $\gamma'' = 0{,}025$. Интервалы такого типа называются центральными [132]. Однако они не являются кратчайшими при одной и той же доверительной вероятности. Более того, в силу несимметричности распределения s, центральные интервалы не могут быть кратчайшими [133].

В табл. 26 приведены кратчайшие доверительные интервалы для σ, равные $z_1 s \leqslant \sigma \leqslant z_2 s$ для заданных доверительных вероятностей. При $n \geqslant 70$ разница между кратчайшими и центральными интервалами становится менее 1%, поэтому кратчайшими интервалами рекомендуется пользоваться при $n \leqslant 70$. Полное сравнительное исследование оценок дисперсии и стандартного отклонения нормального распределения приведено в [134].

Задача 47. *Для данных задачи* 46 (x_i: 1,4; 2,1; 2,9; 3,1; 3,8; 4,1; 4,3; 4,6; 5,1; 6,1) *найти двустороннюю интервальную оценку для стандартного отклонения при доверительной вероятности* $\alpha = 0{,}95$.

Интервальная оценка дисперсии (2.1.2.2.1)

Находим $S^2 = \dfrac{1}{10-1}\sum\limits_{i=1}^{10}(x_i - \bar{x})^2 = 1{,}9872$, или $\sum\limits_{i=1}^{10}(x_i - \bar{x})^2 = 17{,}885$.

Для $\alpha = 0{,}95$ (двусторонние границы) из табл. 22 имеем $u_{0,975} = 1{,}95996$ и $u_{0,025} = -1{,}95996$.

Отсюда при $f = n - 1 = 9$

$$\chi^2_{0,975} = 9 \cdot \left[1 - \frac{2}{9\cdot 10} + 1{,}95996 \cdot \sqrt{\frac{2}{9\cdot 10}}\right]^3 = 18{,}4333;$$

$$\chi^2_{0,025} = 9 \cdot \left[1 - \frac{2}{9\cdot 10} - 1{,}95996 \cdot \sqrt{\frac{2}{9\cdot 10}}\right]^3 = 2{,}9004.$$

Далее

$$(s_n^2)^{\text{н}} = \frac{1}{18{,}4333}\cdot 17{,}885 = 0{,}97025; \quad (s_n^2)^{\text{в}} = \frac{1}{2{,}9004}\cdot 17{,}885 = 6{,}1664.$$

Следовательно, с вероятностью 0,95 имеем $0{,}97025 \leqslant \sigma^2 \leqslant 6{,}1664$.

Интервальная оценка σ по размаху (2.1.2.2.2)

Имеем $x_{\max} - x_{\min} = 6{,}1 - 1{,}4 = 4{,}7$. Из табл. 23 для $n = 10$ и $\alpha = 0{,}95$ имеем $\omega_{10}(0{,}975) = 4{,}79$ и $\omega_{10}(0{,}025) = 1{,}67$. Следовательно,

$$s_n^{\text{н}} = \frac{4{,}7}{4{,}97} = 0{,}9457; \quad s_n^{\text{в}} = \frac{4{,}7}{1{,}67} = 2{,}8144,$$

и с вероятностью 0,95 имеем $0{,}9457 \leqslant \sigma \leqslant 2{,}8144$.

Таблица 26

Коэффициенты z_1 и z_2 кратчайших доверительных интервалов [133]

n	$\alpha = 0{,}90$		$\alpha = 0{,}95$		$\alpha = 0{,}99$	
	z_1	z_2	z_1	z_2	z_1	z_2
2	0,272	7,944	0,246	16,004	0,220	80,062
3	0,400	3,112	0,364	4,434	0,310	9,990
4	0,478	2,304	0,439	2,949	0,378	5,131
5	0,531	1,980	0,491	2,405	0,428	3,691
6	0,571	1,805	0,531	2,124	0,467	3,027
7	0,601	1,693	0,562	1,950	0,497	2,647
8	0,627	1,616	0,587	1,832	0,523	2,402
9	0,647	1,558	0,609	1,746	0,544	2,229
10	0,664	1,512	0,627	1,680	0,563	2,101
11	0,679	1,476	0,642	1,628	0,580	2,002
12	0,692	1,447	0,656	1,585	0,594	1,982
13	0,704	1,421	0,668	1,550	0,607	1,858
14	0,714	1,400	0,679	1,520	0,619	1,804
15	0,724	1,381	0,689	1,494	0,629	1,758
16	0,732	1,365	0,698	1,461	0,639	1,718
17	0,740	1,350	0,706	1,451	0,648	1,683
18	0,747	1,337	0,714	1,433	0,656	1,653
19	0,753	1,326	0,721	1,417	0,664	1,625
20	0,759	1,315	0,727	1,402	0,671	1,601
21	0,765	1,306	0,733	1,389	0,678	1,579
22	0,770	1,297	0,739	1,377	0,684	1,559
23	0,775	1,288	0,744	1,366	0,690	1,541
24	0,779	1,281	0,749	1,356	0,695	1,525
25	0,783	1,274	0,754	1,347	0,701	1,509
26	0,788	1,267	0,758	1,338	0,705	1,495
27	0,791	1,261	0,762	1,330	0,710	1,482
28	0,795	1,256	0,766	1,322	0,715	1,470
29	0,798	1,250	0,770	1,315	0,719	1,458
30	0,802	1,245	0,773	1,308	0,723	1,448
35	0,816	1,224	0,789	1,280	0,741	1,403
40	0,827	1,207	0,801	1,258	0,755	1,369
45	0,837	1,193	0,812	1,241	0,767	1,342
50	0,844	1,182	0,821	1,226	0,778	1,320
60	0,857	1,164	0,835	1,203	0,795	1,285
70	0,868	1,151	0,847	1,186	0,808	1,260

Оценка σ по среднему абсолютному отклонению (2.1.2.2.3)

Находим $\sum\limits_{i=1}^{10} |x_i - \bar{x}| = \sum\limits_{i=1}^{10} |x_i - 3{,}75| = 11$. Из табл. 24 для $\alpha = 0{,}95$ и $n = 10$ имеем $m_{0,975} = 1{,}156$ и $m_{0,025} = 0{,}47$, тогда

$$s_n^{\text{н}} = \frac{1}{10 \cdot 1{,}156} \cdot 11 = 0{,}9515; \quad s_n^{\text{в}} = \frac{1}{10 \cdot 0{,}47} = 2{,}3404.$$

Окончательно имеем $0{,}9515 \leqslant \sigma \leqslant 2{,}3404$.

Интервальная оценка σ, основанная на ее точечной оценке (2.1.2.2.4)

Имеем $s = \sqrt{\frac{1}{9} \sum\limits_{i=1}^{10} (x_i - \bar{x})^2} = 1{,}40969$. Из табл. 25 для $\alpha = 0{,}95$ и $n - 1 = 9$ имеем

$$\sqrt{\frac{n-1}{\chi^2_{0,975}}} = 0{,}688; \quad \sqrt{\frac{n-1}{\chi^2_{0,025}}} = 1{,}83.$$

Тогда

$$s_n^{\text{н}} = 0{,}688 \cdot 1{,}40969 = 0{,}96986; \quad s_n^{\text{в}} = 1{,}83 \cdot 1{,}40969 = 2{,}57973.$$

Если воспользоваться аппроксимациями, то

$$\sqrt{\frac{n-1}{\chi^2_{0,975}}} = \frac{\sqrt{2 \cdot 10 - 1{,}74}}{1{,}96 + \sqrt{20-2}} = 0{,}6889 \quad \text{и} \quad \sqrt{\frac{n-1}{\chi^2_{0,025}}} = \frac{\sqrt{2 \cdot 10 - 0{,}47}}{\sqrt{20-0{,}8} - 1{,}96} = 1{,}8248,$$

что очень близко к табличным значениям.

Теперь воспользуемся кратчайшим доверительным интервалом. Из табл. 26 имеем для $\alpha = 0{,}95$ и $n = 10$: $z_1 = 0{,}642$ и $z_2 = 1{,}628$, откуда

$$s_n^{\text{н}} = 0{,}642 \cdot 1{,}40969 = 0{,}90502 \quad \text{и} \quad s_n^{\text{в}} = 1{,}628 \cdot 1{,}40969 = 2{,}29497,$$

т. е.

$$0{,}90502 \leqslant \sigma \leqslant 2{,}29497.$$

Видим, что в этом случае длина доверительного интервала равна $2{,}29497 - 0{,}90502 = 1{,}38995$ по сравнению с $2{,}57973 - 0{,}969986 = 1{,}60974$ в случае центрального интервала (т. е. меньше на $\approx 14\%$).

2.1.3. Оценки в усеченных и цензурированных выборках

На практике встречаются ситуации, когда некоторые выборочные значения случайной величины отсутствуют. Например, при испытаниях электронных приборов на гарантийную наработку для части приборов фиксируются значения наработки, а для остальных известно только, что их наработка не меньше некоторой гарантируемой величины.

Выборки, в которых отсутствуют значения случайной величины, бо́льшие (или меньшие) некоторого граничного значения, называются усеченными. Если степень усечения известна заранее, то имеет место так называемая не полностью определенная выборка. Выборки, в которых часть членов отбрасывается, называются цензурированными (например, при измерениях отбрасываются крайние значения, как наиболее грубые).

2.1.3.1. Оценки максимального правдоподобия

2.1.3.1.1. Оценки в усеченных выборках

Оценка параметров усеченного нормального распределения производится по формулам

$$\bar{x} = -zs; \quad s = \frac{1}{n}\sum_{i=1}^{n} x_i g(z),$$

где $z = f(y)$ — функция аргумента $y = \dfrac{n\sum\limits_{i=1}^{n} x_i^2}{2\left(\sum\limits_{i=1}^{n} x_i\right)^2}$; $g(z)$ — функция аргумента z.

Значения функций $z = f(y)$ и $g(z)$ табулированы в [29], часть этих таблиц воспроизведена в табл. 27 и 28.

Сначала по выборочным данным подсчитывается y, затем по табл. 27 и 28 — z и $g(z)$ (при необходимости используется интерполяция). В приведенных формулах предполагается, что точка усечения x_y известна и равна 0 (если $x_y \neq 0$, то формулы справедливы для переменной $x_i^* = x_i - x_y$).

Таблица 27

Значения функции $z = f(y)$ [29]

y	z	y	z	y	z
0,550	−3,145	0,670	−1,209	0,790	0,052
0,555	−2,990	0,675	−1,158	0,795	0,110
0,560	−2,851	0,680	−1,103	0,800	0,168
0,565	−2,777	0,685	−1,051	0,805	0,228
0,570	−2,613	0,690	−0,999	0,810	0,289
0,575	−2,508	0,695	−0,947	0,815	0,351
0,580	−2,410	0,700	−0,896	0,820	0,414
0,585	−2,319	0,705	−0,845	0,825	0,479
0,590	−2,232	0,710	−0,894	0,830	0,545
0,595	−2,151	0,715	−0,743	0,835	0,613
0,600	−2,073	0,720	−0,692	0,840	0,683
0,605	−1,999	0,725	−0,641	0,845	0,754
0,610	−1,928	0,730	−0,589	0,850	0,829
0,615	−1,859	0,735	−0,538	0,855	0,905
0,620	−1,792	0,740	0,487	0,860	0,984
0,625	−1,728	0,745	−0,435	0,865	1,066
0,630	−1,665	0,750	−0,383	0,870	1,151
0,635	−1,604	0,755	−0,330	0,875	1,240
0,640	−1,545	0,760	−0,277	0,880	1,332
0,645	−1,486	0,765	−0,224	0,885	1,428
0,650	−1,429	0,770	−0,170	0,890	1,530
0,655	−1,373	0,775	−0,116	0,895	1,636
0,660	−1,318	0,780	−0,060	0,900	1,749
0,665	−1,263	0,785	−0,040	0,905	1,868

Таблица 28

Значения функции $g(z)$ [29]

z	$g(z)$	z	$g(z)$	z	$g(z)$	z	$g(z)$
−3,0	0,3328	−1,8	0,5314	−0,6	0,9442	0,6	1,6935
−2,9	0,3341	−1,7	0,5560	−0,5	0,9909	0,7	1,7624
−2,8	0,3561	−1,6	0,5823	−0,4	1,0396	0,8	1,8327
−2,7	0,3689	−1,5	0,6102	−0,3	1,0902	0,9	1,9043
−2,6	0,3826	−1,4	0,6398	−0,2	1,1428	1,0	1,9771
−2,5	0,3977	−1,3	0,6713	−0,1	1,1917	1,1	2,0511
−2,4	0,4128	−1,2	0,7045	0,0	1,2533	1,2	2,2024
−2,3	0,4294	−1,1	0,7396	0,1	1,3113	1,3	2,2796
−2,2	0,4472	−1,0	0,7766	0,2	1,4323	1,4	2,3578
−2,1	0,4662	−0,9	0,8156	0,3	1,4953	1,5	2,4369
−2,0	0,4866	−0,8	0,8565	0,4	1,5599	1,6	2,5169
−1,9	0,5082	−0,7	0,8993	0,5	1,6259	1,7	2,5978

2.1.3.1.2. Оценки в неполностью определенных выборках

Неполностью определенная выборка имеет место, когда о части членов выборки известно только, что они не больше (не меньше) некоторого граничного значения.

Предположим, что имеется выборка объема n из нормального распределения. Для $(n - n_0)$ членов выборки известны точные значения x_i $(x_1, x_2, x_3 \ldots, x_{n-n_0})$.

Таблица 29

Значения функции $z = f(h, y)$ [29]

y	h							
	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40
0,50		−2,680	−0,985	−1,537	−1,207	−0,943	−0,720	−0,523
0,51		−2,445	−1,862	−1,463	−1,158	−0,909	−0,695	−0,505
0,52		−2,263	−1,759	−1,398	−1,114	−0,877	−0,672	−0,488
0,53	−2,786	−2,115	−1,671	−1,340	−1,074	−0,849	−0,651	−0,472
0,54	−2,562	−1,992	−1,595	−1,289	−1,037	−0,822	−0,631	−0,457
0,55	−2,384	−1,889	−1,528	−1,242	−1,004	−0,797	−0,612	−0,443
0,56	−2,239	−1,800	−1,468	−1,200	−0,973	−0,774	−0,595	−0,429
0,57	−2,117	−1,722	−1,415	−1,162	−0,945	−0,753	−0,579	−0,416
0,58	−2,013	−1,653	−1,366	−1,127	−0,919	−0,733	−0,563	−0,404
0,59	−1,923	−1,591	−1,322	−1,094	−0,894	−0,714	−0,548	−0,393
0,60	−1,844	−1,536	−1,282	−1,064	−0,871	−0,696	−0,534	−0,382
0,62	−1,710	−1,441	−1,211	−1,010	−0,829	−0,663	−0,508	−0,361
0,64	−1,602	−1,360	−1,150	−0,963	−0,792	−0,634	−0,485	−0,343
0,66	−1,512	−1,292	−1,097	−0,921	−0,759	−0,607	−0,463	−0,325
0,68	−1,435	−1,233	−1,050	−0,884	−0,729	−0,583	−0,444	−0,309
0,70	−1,369	−1,180	−1,009	−0,850	−0,702	−0,561	−0,425	0,294
0,72	−1,311	−1,134	−0,971	−0,819	−0,677	−0,540	−0,408	−0,280
0,74	−1,260	−1,092	−0,937	−0,792	−0,654	−0,521	−0,393	−0,267
0,76	−1,214	−1,055	−0,906	−0,766	−0,632	−0,503	−0,378	−0,255
0,78	−1,173	−1,021	−0,878	−0,742	−0,613	−0,487	−0,364	−0,243
0,80	−1,135	−0,990	−0,852	−0,721	−0,594	−0,471	−0,351	−0,232
0,85	−1,055	−0,922	−0,795	−0,673	−0,553	−0,437	−0,322	−0,207
0,90	−0,989	−0,866	−0,747	−0,631	−0,518	−0,407	−0,296	−0,186
0,95	−0,934	−0,819	−0,706	−0,596	−0,488	−0,380	−0,273	−0,166
1,00	−0,887	−0,778	−0,671	−0,565	−0,461	−0,357	−0,253	−0,149
1,30	−0,702	−0,613	−0,525	−0,436	−0,347	−0,257	−0,165	−0,072
1,50	−0,626	−0,545	−0,463	−0,381	−0,297	−0,212	−0,125	−0,036

y	h							
	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80
0,50	−0,345	−0,178	−0,019	0,136	0,291	0,449	0,616	0,797
0,51	−0,331	−0,168	−0,017	0,142	0,295	0,453	0,619	0,799
0,52	−0,318	−0,158	−0,004	0,147	0,300	0,456	0,621	0,801
0,53	−0,306	−0,149	0,003	0,153	0,304	0,459	0,623	0,802
0,54	−0,295	−0,140	0,010	0,158	0,308	0,462	0,625	0,804
0,55	−0,284	−0,132	0,016	0,163	0,312	0,465	0,628	0,806
0,56	−0,273	−0,124	0,023	0,168	0,315	0,468	0,630	0,807
0,57	−0,263	−0,116	0,029	0,173	0,319	0,471	0,632	0,809
0,58	−0,254	−0,108	0,035	0,177	0,323	0,473	0,634	0,810
0,59	−0,245	−0,101	0,040	0,182	0,326	0,476	0,636	0,812
0,60	−0,236	−0,094	0,046	0,186	0,330	0,479	0,638	0,813
0,62	−0,220	−0,081	0,056	0,194	0,336	0,484	0,642	0,816
0,64	−0,204	−0,069	0,066	0,202	0,342	0,488	0,645	0,818
0,66	−0,190	−0,058	0,075	0,210	0,348	0,493	0,649	0,821
0,68	−0,177	−0,047	0,084	0,216	0,353	0,497	0,652	0,823
0,70	−0,166	−0,037	0,092	0,223	0,359	0,501	0,655	0,826
0,72	−0,154	−0,027	0,100	0,229	0,364	0,505	0,658	0,828
0,74	−0,143	−0,019	0,107	0,235	0,368	0,509	0,661	0,830
0,76	−0,133	−0,010	0,114	0,241	0,373	0,513	0,664	0,833
0,78	−0,123	−0,002	0,121	0,246	0,377	0,516	0,667	0,835
0,80	−0,114	0,006	0,127	0,252	0,382	0,520	0,670	0,837
0,90	−0,074	0,039	0,154	0,275	0,401	0,535	0,681	0,846
1,00	−0,043	0,066	0,177	0,294	0,416	0,548	0,692	0,854
1,30	0,025	0,124	0,237	0,336	0,452	0,578	0,716	0,873
1,50	0,056	0,151	0,251	0,357	0,470	0,593	0,728	0,883

Оценки параметров μ и σ тогда подсчитываются по формулам [29]

$$\bar{x} = -zs; \quad s = \frac{1}{n - n_0} \sum_{i=1}^{n-n_0} x_i \frac{1-h}{h\psi(z) - (1-h)z},$$

где $h = \dfrac{n_0}{n}$; $z = f(h, y)$ — функция аргументов h и $y = (n - n_0) \sum\limits_{i=1}^{n-n_0} \dfrac{x_i^2}{2\left(\sum\limits_{i=1}^{n-n_0} x_i\right)^2}$, — значения табулированы в [29]; $\psi(z)$ — функция аргумента z (значения приведены в [29]).

Таблица 30

Значения функции $\psi(z)$ [29]

z	$\psi(z)$	z	$\psi(z)$	z	$\psi(z)$
−3,0	3,2831	−1,3	1,7704	0,4	0,5619
−2,9	3,1903	−1,2	1,6876	0,5	0,5092
−2,8	3,0979	−1,1	1,6058	0,6	0,4591
−2,7	3,0058	−1,0	1,5251	0,7	0,4119
−2,6	2,9141	−0,9	1,4456	0,8	0,3676
−2,5	1,8227	−0,8	1,3674	0,9	0,3251
−2,4	2,7318	−0,7	1,2905	1,0	0,2876
−2,3	2,6414	−0,6	1,2150	1,1	0,2520
−2,2	2,5515	−0,5	1,1411	1,2	0,2194
−2,1	2,4621	−0,4	1,0688	1,3	0,1897
2,0	2,3732	−0,3	0,9982	1,4	0,1629
−1,9	2,2849	−0,2	0,9294	1,5	0,1388
−1,8	2,1973	−0,1	0,8626	1,6	0,1173
−1,7	2,1103	0,0	0,7979	1,7	0,0984
−1,6	2,0241	0,1	0,7353	1,8	0,0819
−1,5	1,9387	0,2	0,6751	1,9	0,0676
−1,4	1,8541	0,3	0,6172	2,0	0,0552

Значения функций $z = f(h, y)$ и $\psi(z)$ воспроизведены в табл. 29 и 30. Сначала по выборочным значениям x_i вычисляются y и h, затем по табл. 29 и 30 определяются соответствующие им значения $z = f(h, y)$ и $\psi(z)$.

2.1.3.1.3. Оценки в цензурированных выборках

2.1.3.1.3.1. Оценка максимального правдоподобия

Наиболее полно оценки такого типа рассмотрены в [135, 136]. Пусть из выборки объема n известны только r первых членов. Это, кстати, классическая ситуация, возникающая при проведении испытаний приборов на долговечность, когда из n испытываемых приборов наблюдаются r отказов.

Итак, наблюдаются значения $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_r$, а для всех оставшихся $(n - r)$ приборов примем одно значение x_0, равное x_r.

Для первых r членов выборки имеем

$$\bar{x} = \frac{1}{r} \sum_{i=1}^{r} x_i \quad \text{и} \quad s^2 = \frac{1}{r} \sum_{i=1}^{r} (x_i - \bar{x})^2.$$

Далее находим параметры $h = \dfrac{n-r}{n}$ и $\gamma = \dfrac{s^2}{(x_0 - \bar{x})^2}$.

Оценки для μ и σ будут иметь вид [136]

$$\hat{\mu} = \bar{x} + (x_0 - \bar{x})k(h,\gamma); \quad \hat{\sigma} = [s^2 + (x_0 - \bar{x})^2 k(h,\gamma)]^{\frac{1}{2}},$$

где $k(h,\gamma)$ — коэффициенты, приведенные в табл. 31.

Таблица 31

Значения $k(h,\gamma)$ [136]

γ	h								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,0	0,1102	0,2426	0,4021	0,5961	0,8368	1,1450	1,5610	2,1760	3,2830
0,1	0,1180	0,2574	0,4233	0,6234	0,8703	1,1850	1,6080	2,2290	3,3450
0,2	0,1247	0,2703	0,4422	0,6483	0,9012	1,2220	1,6510	2,2800	3,4050
0,3	0,1306	0,2819	0,4595	0,6713	0,9300	1,2570	1,6930	2,3290	3,4640
0,4	0,1386	0,2926	0,4755	0,6927	0,9570	1,2900	1,7320	2,3760	3,5200
0,5	0,1409	0,3025	0,4904	0,7129	0,9826	1,3210	1,7700	2,4210	3,5750
0,6	0,1455	0,3118	0,5045	0,7320	1,0070	1,3510	1,8060	2,4650	3,6280
0,7	0,1499	0,3207	0,5180	0,7502	1,0300	1,3800	1,8410	2,5070	3,6790
0,8	0,1540	0,3290	0,5308	0,7676	1,0530	1,4080	1,8750	2,5480	3,7300
0,9	0,1579	0,3370	0,5340	0,7844	1,0740	1,4350	1,9080	2,5880	3,7790
1,0	0,1617	0,3447	0,5548	0,8005	1,0950	1,4610	1,9400	2,6260	3,8270
2,0	0,1932	0,4093	0,6547	0,9382	1,2740	1,6860	2,2170	2,9680	4,2580
3,0	0,2182	0,4609	0,7349	1,0490	1,4200	1,8700	2,4470	3,2550	4,6250
4,0	0,2395	0,5052	0,8038	1,1460	1,5460	2,0310	2,6490	3,5080	4,9520
5,0	0,2585	0,5450	0,8653	1,2310	1,6590	2,1750	2,8290	3,7360	5,2490
6,0	0,2757	0,5803	0,8912	1,3090	1,7620	2,3070	2,9950	3,9450	5,5220
7,0	0,2916	0,6134	0,9729	1,3820	1,8570	2,4280	3,1490	4,1400	5,7880
8,0	0,3065	0,6442	1,0210	1,4400	1,9460	2,5430	3,2930	4,3220	6,0180
9,0	0,3205	0,6733	1,0670	1,5130	2,0310	2,6500	3,4300	4,4950	6,2450
10,0	0,3337	0,7009	1,1100	1,5730	2,1100	2,7530	3,5590	4,6600	4,4620

Двусторонние доверительные интервалы с доверительной вероятностью α находятся по формулам [136]

$$\hat{\sigma}_n^{\text{н}}(\alpha) = \hat{\sigma}\chi'(\alpha,n,r); \quad \hat{\sigma}_n^{\text{в}}(\alpha) = \hat{\sigma}\chi''(\alpha,n,r),$$

где $\chi'(\alpha,n,r), \chi''(\alpha,n,r)$ — коэффициенты, приведенные в табл. 32. При $r = n$ (отсутствие цензурирования) оценки превращаются в обычные.

Двусторонний доверительный интервал для μ имеет вид [136]

$$\hat{\mu}_n^{\text{н}}(\alpha) = \hat{\mu} - t'(\alpha,n,r)\hat{\sigma}; \quad \hat{\mu}_n^{\text{в}}(\alpha) = \hat{\mu} - t''(\alpha,n,r)\hat{\sigma},$$

где $t'(\alpha,n,r)$, $t''(\alpha,n,r)$ — коэффициенты, приведенные в табл. 33.

Отметим основные особенности рассмотренных оценок:

— точная длина доверительного интервала в большей степени зависит от r и в меньшей степени от $h = \dfrac{n-r}{n}$;

— при фиксированном значении r длина доверительного интервала для μ остается практически постоянной при $0 \leqslant h \leqslant 0{,}7$ и увеличивается при $h > 0{,}7$;

— при $r = \text{const}$ доверительный интервал для σ медленно увеличивается с ростом n (или h);

— при $n = \text{const}$ длина доверительного интервала уменьшается с ростом r.

Таблица 32

Коэффициенты $\chi'(\alpha, n, r)$ (верхняя строка) **и $\chi''(\alpha, n, r)$** (нижняя строка) [119]

r	n	$\alpha = 0{,}90$	$\alpha = 0{,}95$	r	n	$\alpha = 0{,}90$	$\alpha = 0{,}95$	r	n	$\alpha = 0{,}90$	$\alpha = 0{,}95$
2	4,5	0,7187	0,6097	6	10	0,7194	0,6622	15	20	0,7874	0,7463
		27,7780	58,8230			2,5190	2,9250			1,5620	1,6950
	10	0,7143	0,6024		20	0,7143	0,6536		50	0,7752	0,7299
		31,2500	66,1670			2,6310	3,1250			1,6130	1,7540
	50	0,7092	0,5952	7	8	0,7407	0,6896		100	0,7575	0,7143
		34,4830	66,1670			2,1280	2,4390			1,6670	1,8180
3	4	0,7042	0,6329		10	0,7299	0,6803	20	30	0,8064	0,7633
		5,8820	8,3330			2,2220	2,6310			1,3500	1,5850
	5	0,7042	0,6211		20	0,7246	0,6667		50	0,8000	0,7576
		6,2500	9,0900			2,3250	2,7030			1,4920	1,6130
	20	0,6849	0,6060	8	9	0,7519	0,7042		100	0,7874	0,7463
		7,1430	10,0000			1,9610	2,2220			1,5150	1,6390
4	5	0,7092	0,6452		10	0,7463	0,6944	25	30	0,8196	0,7936
		3,5710	4,5450			2,0410	2,3250			1,3510	1,4490
	10	0,6993	0,6289		20	0,7353	0,6944		50	0,8130	0,7752
		4,1660	5,2630			2,1280	3,4390			1,3890	1,4700
	20	0,6896	0,6173		50	0,7194	0,6666		100	0,8064	0,7692
		4,1660	5,5550			2,2220	2,5640			1,4280	1,5150
5	6	0,7194	0,6622	10	12	0,7633	0,7143	30	50	0,8624	0,7692
		2,7780	3,4480			1,7860	2,0000			1,3330	1,4080
	10	0,7092	0,6452		20	0,7519	0,7042		100	0,8817	0,7874
		3,0300	3,7100			1,8520	2,0830			1,3700	1,4490
	20	0,7042	0,6329		50	0,7407	0,6896	50	70	0,8620	0,8333
		3,1250	4,0000			1,9230	2,1740			1,2340	1,2820
	50	0,6896	0,6250		100	0,7799	0,6803		100	0,8547	0,8264
		3,2260	4,0000			2,0000	2,2220			1,2500	1,2990
	100	0,6803	0,6135	12	15	0,7752	0,7299	70	100	0,8771	0,8547
		3,3330	4,1660			1,6660	1,8180			1,1900	1,2190
6	7	0,7299	0,6803		20	0,7692	0,7246	90	100	0,8928	0,8772
		2,3810	2,7780			1,6950	1,8870			1,1490	1,1760
					50	0,7519	0,7042				
						1,7540	1,9610				

2.1.3.1.3.2. Оценки с помощью порядковых статистик

Предположим, что цензурирование заключается в отбрасывании из выборки r_1 наименьших и r_2 наибольших членов, а параметры μ и σ оцениваются по оставшимся $(n - r_1 - r_2)$ наблюдениям.

Наиболее просты в вычислительном отношении оптимальные линейные оценки [118, 119]. Они находятся по формулам

$$\mu_n = \sum_{i=r_1+1}^{n-r_2} k_i x_i; \quad \sigma_n = \sum_{i=r_1+1}^{n-r_2} k_i' x_i,$$

где x_i — i-я порядковая статистика в выборке, упорядоченной по убыванию $(x_1 \geqslant x_2 \geqslant \ldots \geqslant x_n)$; k_i, k_i' — коэффициенты оценки, табулированные в [118, 119].

Коэффициенты k_i и k_i' для выборки объема $n = 10$ приведены в табл. 34.

Очевидно, что таблицы для k_i и k_i' при различных сочетаниях n, $r_1(r_2)$ очень громоздки и практически неприменимы. Поэтому рекомендуется использовать так

Таблица 33

Коэффициенты $t'(\alpha, n, r)$ (верхняя строка) **и $(-1) \cdot t''(\alpha, n, r)$** (нижняя строка) [119]

r	n	α		r	n	α		r	n	α		r	n	α	
		0,90	0,95			0,90	0,95			0,90	0,95			0,90	0,95
2	5	1,08	1,75	4	12	0,52	0,64	7	15	0,46	0,56	15	90	0,33	0,40
		19,60	41,20			2,78	3,86			1,09	1,40			0,85	1,05
	6	0,79	1,22		15	0,48	0,53		20	0,41	0,49		100	0,34	0,40
		23,90	51,70			3,29	4,61			1,25	1,63			0,89	1,12
	7	0,65	0,90		20	0,47	0,57	8	10	0,61	0,77	20	40	0,28	0,34
		26,40	55,10			3,68	5,10			0,74	0,94			0,42	0,52
	8	0,60	0,79	5	6	0,90	1,16		12	0,53	0,66		50	0,27	0,32
		30,50	60,60			1,17	1,54			0,80	1,12			0,47	0,59
	10	0,53	0,68		7	0,80	1,03		15	0,46	0,56		70	0,27	0,32
		34,30	70,60			1,26	1,67			0,88	1,13			0,54	0,66
	12	0,50	0,66		8	0,69	0,90		20	0,41	0,49		90	0,28	0,34
		37,80	75,60			1,35	1,85			1,03	1,35			0,59	0,73
	15	0,50	0,65		10	0,59	0,75		30	0,38	0,45		100	0,29	0,34
		40,30	85,80			1,56	2,11			1,25	1,63			0,61	0,76
	20	0,48	0,65		12	0,51	0,63		50	0,40	0,47	25	35	0,29	0,35
		48,20	100,00			1,72	2,33			1,56	2,00			0,34	0,41
	30	0,49	0,70		20	0,44	0,53	10	12	0,53	0,67		50	0,25	0,36
		54,80	112,00			2,31	3,19			0,62	0,78			0,37	0,45
	50	0,56	0,78		30	0,46	0,55		15	0,47	0,57		70	0,24	0,28
		64,10	123,00			2,80	3,78			0,66	0,83			0,42	0,51
3	5	1,10	1,55		50	0,50	0,59		20	0,39	0,47		90	0,25	0,29
		3,38	5,11			3,40	4,62			0,75	0,95			0,46	0,57
	6	0,89	1,25		70	0,52	0,64		30	0,36	0,42		100	0,25	0,30
		3,71	5,61			3,84	5,21			0,89	1,12			0,48	0,59
	7	0,75	1,01		90	0,57	0,69		50	0,36	0,43	30	50	0,24	0,28
		4,31	6,51			4,11	5,47			1,11	1,41			0,31	0,38
	8	0,65	0,85		100	0,58	0,70		70	0,39	0,47		70	0,23	0,27
		4,58	7,10			4,21	5,66			1,30	1,65			0,34	0,42
	10	0,57	0,71	6	7	0,80	1,02		90	0,41	0,49		90	0,23	0,27
		5,55	8,21			0,94	1,21			1,44	1,80			0,37	0,45
	12	0,51	0,64		8	0,70	0,89		100	0,42	0,50		100	0,24	0,28
		6,43	9,82			0,98	1,28			1,45	1,90			0,39	0,47
	15	0,49	0,60		10	0,60	0,75	12	15	0,46	0,57	40	50	0,24	0,28
		7,05	10,70			1,11	1,46			0,55	0,68			0,26	0,32
	20	0,50	0,61		12	0,52	0,65		20	0,39	0,47		70	0,21	0,25
		7,92	1,27			1,20	1,58			0,60	0,75			0,27	0,32
4	5	1,12	1,47		15	0,46	0,56		30	0,33	0,40		90	0,20	0,24
		1,56	2,16			1,42	1,86			0,68	0,86			0,28	0,34
	6	0,91	1,20		20	0,43	0,51	15	25	0,34	0,41		100	0,20	0,24
		1,81	2,51			1,65	2,17			0,49	0,62			0,29	0,35
	7	0,79	1,03	7	8	0,71	0,88		35	0,31	0,37	50	70	0,20	0,24
		1,95	1,95			0,81	1,04			0,55	0,69			0,23	0,27
	8	0,68	0,88		10	0,61	0,76		50	0,38	0,35		90	0,19	0,22
		2,21	3,01			0,88	1,12			0,66	0,81			0,23	0,28
	10	0,58	0,73		12	0,53	0,66		70	0,32	0,37		100	0,18	0,22
		2,60	3,59			0,96	1,12			0,75	0,94			0,24	0,29

Таблица 34

Коэффициенты k_i (верхняя строка) и k'_i (нижняя строка) для $n = 10$

r_1	r_2	i									
		1	2	3	4	5	6	7	8	9	10
0	0	0,1000	0,1000	0,1000	0,1000	0,1000	0,1000	0,1000	0,1000	0,1000	0,1000
		−0,2044	−0,1172	−0,0763	−0,0436	−0,0142	0,0142	0,0436	0,0763	0,1172	0,2044
	1	0,0843	0,0921	0,0957	0,0986	0,1011	0,1036	0,1060	0,1085	0,2101	
		−0,2364	−0,1314	−0,0851	−0,0465	−0,0110	0,0215	0,0559	0,0937	0,3423	
	2	0,0605	0,0804	0,0898	0,0972	0,1037	0,1099	0,1161	0,3424		
		−0,2753	−0,1523	−0,0947	−0,0488	−0,0077	0,0319	0,0722	0,4746		
	3	0,0244	0,0636	0,0818	0,0962	0,1089	0,1207	0,5045			
		−0,3252	−0,1758	−0,1058	−0,0502	−0,0006	0,0469	0,6107			
	4	−0,0316	0,0383	0,0707	0,0962	0,1185	0,7078				
		−0,3930	−0,2063	−0,1192	−0,0501	0,0111	0,7576				
	5	−0,1240	−0,0016	0,0549	0,0990	0,9718					
		−0,4919	−0,2491	−0,1362	−0,0472	0,9243					
	6	−0,2923	−0,0709	0,0305	1,3327						
		−0,6520	−0,3150	−0,1593	1,1263						
	7	−0,6598	−0,2138	1,8734							
		−0,9625	−0,4357	1,3981							
	8	−1,8634	2,8634								
		−1,8608	1,8608				0,1041				
1	1		0,1884	0,1036	0,1040	0,1041	0,0201	0,1040	0,1036	0,1884	
			−0,4034	−0,1074	−0,0616	−0,0201	0,1098	0,0616	0,1074	0,4034	
	2		0,7525	0,0961	0,1013	0,1057	0,0325	0,1138	0,3209		
			−0,4803	−0,1235	−0,0674	−0,0166	0,1204	0,0827	0,0572		
	3		0,0942	0,0846	0,0979	0,1095	0,0514	0,4933			
			−0,5842	−0,1440	−0,0734	−0,0097	0,7261	0,7599			
	4		−0,0043	0,0665	0,0938	0,1179	0,9844				
			−0,7959	−0,1719	−0,0797	0,0031					
	5		−0,1866	0,0351	0,0892	1,0623					
			−1,0120	−0,2145	−0,0859	1,2835					
	6	−0,5877	−0,2918	1,6166							
		−1,4678	2,9000	1,7595							
	7	−1,9000	2,8960								
		−2,8960					0,1103				
2	2			0,2798	0,1099	0,1103	0,0310	0,1099	0,2798		
				−0,7021	−0,0947	−0,0310	0,1198	0,0947	0,7021		
	3			0,2050	0,1038	0,1122	0,0549	0,4592			
				−0,8898	−0,1101	−0,0262	0,7281	0,9711			
	4			0,0706	0,0935	0,1178	1,3415				
				−1,1952	−0,1318	−0,0144					
	5			−0,2648	0,0735	1,1914					
				−1,7947	−0,1688	1,9635					
	6			−1,3406	2,3406						
				−3,5677	3,5677		0,1193				
3	3				0,3807	0,1193	0,0559	0,3807			
					−1,2832	−0,0589	0,6930	1,2832			
	4				0,1871	0,1198	2,0344				
					−1,9791	0,0553					
	5				−0,4747	1,4847					
					−3,9511	3,9511	0,5000				
4	4					0,5000	4,0761				
						−4,0761					

называемые альтернативные оценки Гупты [137, 138], вычисляемые по формулам

$$\tilde{\mu}_n = \sum_{i=r_1+1}^{n-r_2} a_i x_i; \quad \tilde{\sigma}_n = \sum_{i=r_1+1}^{n-r_2} b_i x_i,$$

где a_i и b_i — весовые коэффициенты, вычисляемые по формулам

$$a_i = \frac{1}{n-r_1-r_2} - \frac{\bar{U}(U_i-\bar{U})}{\sum\limits_{i=r_1+1}^{n-r_2}(U_i-\bar{U})^2}; \quad b_i = \frac{U_i-\bar{U}}{\sum\limits_{i=r_1+1}^{n-r_2}(U_i-\bar{U})^2};$$

U_i — математическое ожидание i-й порядковой статистики из стандартного нормального распределения;

$$\bar{U} = \frac{1}{n-r_1-r_2}\sum_{i=r_1+1}^{n-r_2} U_i.$$

Эффективность этой оценки составляет $0{,}96 \div 0{,}98$ от оптимальной линейной при $n > 10$. Напомним, что для U_i может быть использована полезная аппроксимация

$$U_i = 4{,}91(n+0{,}25)^{-0{,}14}\left[(i-0{,}375)^{0{,}14} - (n-i+0{,}625)^{0{,}14}\right].$$

В заключение приведем простую оценку для μ, предложенную Диксоном [139]:

$$\tilde{\tilde{\mu}}_n = \frac{1}{n}\sum_{i=2}^{n-2r-1} x_{r+i} + \frac{(r+1)(x_{r+1}+x_{n-r})}{n},$$

где $r = \max(r_1, r_2)$.

Эффективность этой оценки не уступает эффективности оптимальной линейной оценки.

Задача 48. *Испытаны 15 приборов. При этом значения параметра — критерия годности зафиксированы у 10 приборов:*

$$x_i: 1{,}1; 2{,}1; 2{,}4; 3{,}1; 3{,}5; 3{,}7; 4{,}2; 4{,}8; 5{,}9; 6{,}3,$$

а для остальных известно, что их наработка имеет большее значение. Необходимо вычислить оценки максимального правдоподобия для μ и σ.

Находим $\sum\limits_{i=1}^{10} x_i = 37{,}1$; $\sum\limits_{i=1}^{10} x_i^2 = 162{,}11$ и $y = \dfrac{10\cdot 162{,}11}{2\cdot(37{,}1)^2} \approx 0{,}59$.

Для $y = 0{,}59$ из табл. 27 имеем $z = -2{,}232$, а из табл. 28 (интерполируя) имеем $g(z) \approx 0{,}439$. Окончательно оценки равны

$$s = \sum_{i=1}^{10}\frac{x_i}{10}\cdot g(z) = \frac{0{,}439}{10}\cdot\sum_{i=1}^{10} x_i = 0{,}0439\cdot 37{,}1 = 1{,}628; \quad \bar{x} = 2{,}232\cdot 1{,}628 = 3{,}635.$$

Задача 49. *Решить задачу 48, исходя из того, что степень усечения известна (предположим, она равна оценке $h = n_0/n$).*

Пользуемся формулами для неполностью определенной выборки (см. раздел 2.1.3.1.2). Имеем $y = \dfrac{10\sum\limits_{i=1}^{10} x_i^2}{2\left(\sum\limits_{i=1}^{10} x_i\right)^2} = 0{,}59$.

Оценку нормированной точки усечения $z = f(h, y)$ находим из табл. 29 ($h = 0{,}333$ и $y = 0{,}59$): $z = f(0{,}333; 0{,}59) = -0{,}631$.

Далее из табл. 30 имеем $\psi(z) = 0{,}4355$ и находим

$$\frac{1-h}{h \cdot \psi(z) - (1-h) \cdot z} = \frac{1 - 0{,}333}{0{,}333 \cdot 0{,}4355 + 0{,}6666 \cdot 0{,}631} = 1{,}17918.$$

Окончательно, $s = \frac{1}{10} \cdot 37{,}1 \cdot 1{,}17918 = 4{,}374$; $\bar{x} = 0{,}631 \cdot 4{,}374 = 2{,}756$.

Разница в оценках является следствием отклонения истинной степени усечения от ее оценки, полученной по малой выборке.

Задача 50. *Для данных задачи 48 найти оценки μ и σ, исходя из того, что в выборке объема $n = 10$ два наибольших наблюдения цензурированы.*

Из условия задачи следует, что известны 8 значений:

$$x_i: 1{,}1; 2{,}1; 2{,}4; 3{,}1; ; 3{,}53{,}7; 4{,}2; 4{,}8,$$

а два наибольших члена выборки (5,9 и 6,3) из нее исключены (цензурированы).

Оценка максимального правдоподобия (2.1.3.1.3.1)

Имеем $n = 10$, $r = 8$ и $\bar{x} = \frac{1}{8} \cdot \sum_{i=1}^{8} x_i = \frac{24{,}5}{8} = 3{,}0625$; $s^2 = \frac{1}{8} \cdot \sum_{i=1}^{8} (x_i - \bar{x})^2 = 1{,}3523$.

Далее $h = \frac{10-8}{10} = 0{,}2$ и $\gamma = \frac{s^2}{(x_0 - \bar{x})^2} = \frac{1{,}3523}{(4{,}8 - 3{,}0625)^2} = 0{,}4479$.

Из табл. 31 для $\gamma \approx 0{,}45$ и $h = 0{,}2$ имеем $k(0{,}2; 0{,}45) \approx 0{,}2872$.

Тогда окончательно

$$\hat{\mu} = 3{,}0625 + (4{,}8 - 3{,}0625) \cdot 0{,}2875 = 0{,}5615;$$

$$\hat{\sigma} = [1{,}3523 + (4{,}8 - 3{,}0625)^2 \cdot 0{,}2872]^{\frac{1}{2}} = 1{,}4897.$$

Найдем теперь двусторонний доверительный интервал при $\alpha = 0{,}95$ для μ и σ. Из табл. 32 находим значения $\chi'(0{,}95; 10{,}8) = 0{,}6944$ и $\chi''(0{,}95; 10{,}8) = 2{,}325$ и далее вычисляем:

$$\hat{\sigma}_n^{\text{н}} = \hat{\sigma} \cdot \chi' = 1{,}4897 \cdot 0{,}6944 = 1{,}0344; \quad \hat{\sigma}_n^{\text{в}} = \hat{\sigma} \cdot \chi'' = 1{,}4897 \cdot 2{,}325 = 3{,}4635,$$

т. е. $1{,}0344 \leqslant \hat{\sigma} \leqslant 3{,}4635$ с доверительной вероятностью 0,95.

Теперь из табл. 33 имеем $t'(0{,}95; 10{,}8) = 0{,}77$ и $t''(0{,}95; 10{,}8) = 0{,}94$.

Следовательно,

$$\hat{\mu}_n^{\text{н}} = 3{,}5615 - 0{,}77 \cdot 1{,}4897 = 2{,}414; \hat{\mu}_n^{\text{в}} = 3{,}5615 + 0{,}94 \cdot 1{,}4897 = 4{,}962,$$

и доверительный интервал равен $2{,}414 \leqslant \mu \leqslant 4{,}692$ с вероятностью $\alpha = 0{,}95$.

Оценка с помощью порядковых статистик (2.1.3.1.3.2)

В нашем случае $r_i = 0$; $r_2 = 2$ и $n = 10$. Из табл. 34 находим

$$\begin{array}{llll}
k_1 = 0{,}0605; & k_2 = 0{,}0804; & k_3 = 0{,}0898; & k_4 = 0{,}0972; \\
k_5 = 0{,}1037; & k_6 = 0{,}1099; & k_7 = 0{,}1161; & k_8 = 0{,}3224; \\
k'_1 = -0{,}2753; & k'_2 = -0{,}1523; & k'_3 = -0{,}0947; & k'_4 = -0{,}0488; \\
k'_5 = -0{,}0077; & k'_6 = 0{,}0319; & k'_7 = 0{,}0722; & k_8 = 0{,}4746.
\end{array}$$

Далее имеем

$$\mu_n = \sum_{i=1}^{8} k_i \cdot x_i = 0{,}0605 \cdot 1{,}1 + 0{,}0804 \cdot 2{,}1 + 0{,}0898 \cdot 2{,}4 + 0{,}0972 \cdot 3{,}1 + 0{,}1037 \cdot 3{,}5 + \\ + 0{,}1099 \cdot 3{,}7 + 0{,}1161 \cdot 4{,}2 + 0{,}3424 \cdot 4{,}8 = 3{,}6529;$$

$$\sigma_n = -0{,}2753 \cdot 1{,}1 - 0{,}1523 \cdot 2{,}1 - 0{,}0947 \cdot 2{,}4 - 0{,}0488 \cdot 3{,}1 - 0{,}0077 \cdot 3{,}5 + \\ + 0{,}0319 \cdot 3{,}7 + 0{,}0722 \cdot 4{,}2 + 0{,}4746 \cdot 4{,}8 = 1{,}671.$$

Рассмотрим *простую оценку Диксона* для μ $(r = 2)$

$$\mu_n = \frac{1}{10} \cdot \sum_{\imath=2}^{n-2r-1} x_{r+i} + \frac{(r+1)\cdot(x_{r+1} + x_{n-r})}{n} = \frac{1}{10} \cdot \sum_{\imath=2}^{5} x_{2+\imath} + 3 \cdot \frac{x_3 + x_8}{10} = 3{,}61.$$

В заключение рассмотрим вычисление необходимых оценок с помощью оценок Гупты. Предварительно вычислим математические ожидания необходимых порядковых статистик:

$$U_1 = 4{,}91 \cdot (10{,}25)^{0{,}14} \cdot [(1 - 0{,}375)^{0{,}14} - (n - 1 + 0{,}625)^{0{,}14}] =$$
$$= 3{,}5447{\cdot}[(1 - 0{,}375)^{0{,}14} - (10 - 1 + 0{,}625)^{0{,}14}] = 3{,}5447{\cdot}(0{,}625^{0{,}14} - 9{,}625^{0{,}14}) = -1{,}54797;$$

$$U_2 = -0{,}99874; \quad U_3 = -0{,}65329; \quad U_4 = -0{,}37392; \quad U_5 = 0{,}122032;$$
$$U_6 = 0{,}122033; \quad U_7 = 0{,}37392; \quad U_8 = 0{,}65329.$$

Далее вычисляем:

$$\bar{U} = -0{,}3833875; \quad \sum_{\imath=1}^{8} (U_\imath - \bar{U})^2 = 29{,}967746.$$

Вычисляем коэффициенты оценки:

$$a_1 = \frac{1}{8} - \frac{(-0{,}31833875)\cdot(-1{,}54797 + 0{,}31833875)}{29{,}967746} = 0{,}11194;$$
$$a_2 = 0{,}11772293; \quad a_3 = 0{,}1214419; \quad a_4 = 0{,}1244092; \quad a_5 = 0{,}127085;$$
$$a_6 = 0{,}1296779; \quad a_7 = 0{,}1323536; \quad a_8 = 0{,}1353213.$$

Таким образом, оценка равна

$$\mu_n = 0{,}11194 \cdot 1{,}1 + 0{,}11777 \cdot 2{,}1 + 0{,}12144 \cdot 2{,}4 + 0{,}124409 \cdot 3{,}1 + 0{,}127085 \cdot 3{,}5 +$$
$$+ 0{,}12968 \cdot 3{,}7 + 0{,}132354 \cdot 4{,}2 + 0{,}13532 \cdot 4{,}8 = 3{,}17746.$$

2.2. Оценка параметров экспоненциального распределения

Напомним, что плотность экспоненциального распределения вероятностей случайной величины описывается формулой

$$f(x;\nu) = \frac{1}{\nu}\exp\Big(-\frac{x}{\nu}\Big), \quad x \geqslant 0,$$

где ν — параметр распределения.

Экспоненциальное распределение широко применяется при анализе надежности технических устройств. Поэтому представляет интерес оценка параметра экспоненциального распределения применительно к различным планам испытаний на надежность. В качестве оцениваемого параметра в этом случае рассматривается, как принято в теории надежности, интенсивность отказов $\lambda = \dfrac{1}{\nu}$.

Для обозначения планов испытаний применяется 3-позиционный код, предложенный авторами работы [140]. Первая позиция кода обозначает число испытываемых изделий (объем выборки). Вторая позиция — буква, указывающая, заменяются ли при испытаниях отказавшие приборы или нет (B — заменяются; $Б$ — не заменяются). Третья позиция кодирует условия проведения испытаний (T — испытания ведутся в течение заданного времени T; r — испытания ведутся до получения r отказов). Например, код $[15, Б, 2]$ означает, что оценка λ проводится по результатам испытаний 15 приборов, без замены отказавших в процессе испытаний, до получения двух отказов. Иногда используются смешанные планы. Например, $[N, B, (r, T)]$ — план, при котором N приборов испытываются с заменой отказавших до появления r отказов, но не более времени T.

2.2.1. Точечные оценки

2.2.1.1. Оценка максимального правдоподобия

Обозначим через d число изделий, отказавших за время проведения испытаний T. Тогда оценки параметра λ находятся по формулам [140]:

$$\lambda_{N,B,T} = \frac{d}{NT}; \quad \lambda_{N,B,r} = \frac{r-1}{Nt_r}; \quad \lambda_{N,Б,T} = \frac{d}{\sum\limits_{i=1}^{d} t_i + (N-d)T};$$

$$\lambda_{N,B,(r,T)} = \begin{cases} \dfrac{d}{NT}, & \text{если } t_r > T; \\[2mm] \dfrac{r-1}{Nt_r}, & \text{если } t_r \leqslant T, \end{cases}$$

где t_r — время наступления r-го отказа; t_i — наработка i-го прибора до отказа.

Справедливы аппроксимации

$$\lambda_{N,Б,T} \approx \begin{cases} \dfrac{d}{\Big(N - \dfrac{d}{2}\Big)T} & \text{при } \dfrac{d}{N} \leqslant 0{,}1 \text{ и } d \geqslant 10; \\[4mm] \dfrac{1}{T}\ln\dfrac{N}{N-d} & \text{при } 0{,}2 \leqslant \dfrac{d}{N} \leqslant 0{,}8. \end{cases}$$

Далее

$$\lambda_{N,Б,r} = \frac{r-1}{\sum_{i=1}^{r} t_i + (N-r)t_r}; \quad \lambda_{N,з,(r,T)} = \begin{cases} \dfrac{d}{\sum_{i=1}^{d} t_i + (N-d)\,t_r}, & \text{если } t_r > T; \\ \dfrac{r-1}{\sum_{i=1}^{r} t_i + (N-r)t_r}, & \text{если } t_r \leqslant T. \end{cases}$$

2.2.1.2. Уточненная двухстадийная оценка

В [141] предлагается эффективная оценка параметра ν, основанная на двухстадийной процедуре. Оценка исходит из того факта, что величина $\dfrac{2n\bar{x}}{\nu}$ имеет распределение χ^2 с $f = 2n$ степенями свободы (n — объем выборки, $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$).
Суть процедуры оценки заключается в следующем. На первой стадии рассматривается предполагаемое значение ν_0. Далее, устанавливая приемлемый уровень достоверности α, проверяем справедливость неравенства

$$\chi^2\left(\frac{1-\alpha}{2}\right) < \frac{2n\bar{x}}{\nu_0} < \chi^2\left(\frac{1+\alpha}{2}\right),$$

где $\chi^2(\beta)$ – β-квантиль распределения χ^2 с $f = 2n$ степенями свободы.

Если неравенство не отклоняется, то принимаются более точные оценки

$$\hat{\nu}_1 = \frac{2n\bar{x}}{\nu_0\chi^2\left(\dfrac{1+\alpha}{2}\right)}\bar{x} + \nu_0\left(1 - \frac{2n\bar{x}}{\nu_0\chi^2\left(\dfrac{1+\alpha}{2}\right)}\right)$$

или

$$\hat{\nu}_2 = \left(\frac{2n\bar{x}}{\nu_0\chi^2\left(\dfrac{1+\alpha}{2}\right)}\right)^2\bar{x} + \left(1 - \frac{4n^2\bar{x}^2}{\nu_0^2\left[\chi^2\left(\dfrac{1+\alpha}{2}\right)\right]^2}\right)\nu_0.$$

Эффективность первой оценки в 2÷2,5 раз выше обычной, второй — в 2÷3,5 раз (обычная оценка $\hat{\nu} = \bar{x}$).

2.2.1.3. Оценки, основанные на порядковых статистиках

2.2.1.3.1. Оптимальная линейная оценка

Предположим, что имеет место двустороннее цензурирование, т. е. неизвестны значения r_1 наименьших и r_2 наибольших членов выборки объема n. При $r_1 = r_2 = 0$ будем иметь случай полной (нецензурированной) выборки.

Оптимальная линейная оценка находится по формуле [119]

$$\nu_0 = \frac{1}{c}\left[\left(\frac{m}{l} - n + r_1\right)x_{r_1+1} + r_2 x_{n-r_2} + \sum_{i=r_1+1}^{n-r_2} x_i\right],$$

где

$$c = \frac{m^2}{l} + n - r_1 - r_2 - 1; \quad m = \sum_{i=1}^{r_1+1} a_i; \quad l = \sum_{i=1}^{r_1+1} a_i^2; \quad a_i = \frac{1}{n-i+1}.$$

Здесь x_i — i-я порядковая статистика выборки (i-й по величине член выборки, ранжированной по возрастанию $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$); r_1, r_2 — количества цензурированных соответственно наименьших и наибольших наблюдений в выборке; $\sum\limits_{i=1}^{k} a_i$ — математическое ожидание k-й порядковой статистики в выборке объема n из нормированного экспоненциального распределения.

При $r_1 = r_2 = 0$ оценка ν_0 совпадает с оценкой максимального правдоподобия $\nu_0 = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$.

При $r_1 = 0$ имеем $\nu_0 = \dfrac{1}{n-r_2}\left(\sum\limits_{i=1}^{n-r_2} x_i + r_2 x_n - r_2\right)$, что совпадает с оценкой для плана испытаний $[N, Б, r]$ (при $r_2 = N - r$, см. раздел 2.1.1). Иногда в выборке отсутствует (цензурирована) часть центральных (средних по величине) выборочных значений. Если известны только s наименьших и r наибольших членов выборки, то оценка для ν имеет вид [119]

$$\hat{\nu}_0 = \frac{1}{\beta}\left[\sum_{i=1}^{s-1} x_i + \left(n - s + 1 + \sum_{i=s+1}^{n-r+1} a_i x_s + \frac{\sum\limits_{i=s+1}^{n-r+1} a_i}{\sum\limits_{i=s+1}^{n-r+1} a_i^2} - r + 1\right) x_{n-r+1} + \sum_{i=1}^{r-1} x_{n-r+1+i}\right],$$

где

$$\beta = s + r - 1 + \frac{\left(\sum\limits_{i=s+1}^{n-r+1} a_i\right)^2}{\sum\limits_{i=s+1}^{n-r+1} a_i^2}.$$

2.2.1.3.2. Оценка по одной порядковой статистике

Хартер [142] и Эпштейн [119] предложили оценку ν, основанную на одной порядковой статистике x_r (т. е. на одном r-м по величине наблюдении). Оценка находится по формуле

$$\nu_r = \frac{x_r}{\sum\limits_{i=1}^{r} a_i}.$$

Эпштейн показал, что эффективность этой оценки по сравнению с оценкой ν_0 больше 0,96 при $\dfrac{r}{n} \leqslant \dfrac{2}{3}$ и 0,98 при $\dfrac{r}{n} \leqslant \dfrac{1}{2}$.

2.2.1.3.3. Оценка Эпштейна [119]

Если в выборке цензурированы $(s-1)$ наименьших членов, то оценка находится по формуле

$$\hat{\nu} = \frac{1}{n}\left[\left(s \cdot \frac{1}{\sum\limits_{i=1}^{s} a_i} + s - n\right) x_s + \sum_{i=s+1}^{n} x_i\right].$$

При цензурировании $(s-1)$ наименьших и $(r-1)$ наибольших по величине наблюдений оценка Эпштейна имеет вид

$$\hat{\nu} = \frac{1}{n-r+1}\left[\left(s\cdot\frac{1}{\sum\limits_{i=1}^{s} a_i} + s - n\right)x_s + \sum_{i=s+1}^{n-r} x_i + x_{n-r+1}\right].$$

2.2.1.3.4. Оценка Огавы [119]

Оценка Огавы позволяет произвести оценку параметра $\lambda = 1/\nu$ экспоненциального распределения по ограниченному числу порядковых статистик. Оценки рассмотрены в [142–151] и имеют вид $\nu^0 = \sum\limits_{i=1}^{k} b_i x_{[n\varepsilon_i]+1}$, где k — количество порядковых статистик, по которым производится оценка ($[\ldots]$ — целая часть числа); b_i, ε_i — коэффициенты оценки, приведенные в табл. 35.

Двумя наилучшими наблюдениями, по которым при $k=2$ следует находить оценку ν^0, являются:

x_{n-1}, x_n при $4 \geqslant n \geqslant 2$; x_{n-2}, x_n при $7 \geqslant n \geqslant 5$; x_{n-3}, x_n при $11 \geqslant n \geqslant 8$; x_{n-4}, x_n при $15 \geqslant n \geqslant 12$; x_{n-6}, x_{n-1} при $18 \geqslant n \geqslant 16$; x_{n-7}, x_{n-1} при $21 \geqslant n \geqslant 19$.

Задача 51. *Партия изделий объема $N=100$ испытана на надежность с заменой отказавших приборов в течение времени $T=1000$ ч, при этом наблюдались $d=10$ отказов. Найти оценку интенсивности отказов λ.*

Для плана испытаний $[100, B, T=1000]$ имеем

$$\lambda_{N,B,T} = \frac{d}{N\cdot T} = \frac{10}{100\cdot 1000} = 10^{-5}\ \text{ч}^{-1}.$$

Задача 52. *Найти оценку интенсивности отказов в условиях задачи* 51, *приняв, что* 100 *ч — это момент отказа* 10*-го изделия.*

Имеем $r=10$, $t_r = 1000$ ч. Для плана $[100, B, r=10]$ получаем

$$\lambda_{N,B,r} = \frac{r-1}{N\cdot t_r} = \frac{10-1}{100\cdot 1000} = 9\cdot 10^{-6}\ \text{ч}^{-1}.$$

Задача 53. *Партия изделий объема $N=10$ шт. была испытана на надежность в течение* 1000 *ч без замены отказавших изделий. При этом были зафиксированы* 5 *отказов в моменты времени (ч)*

$$t_1 = 120,\quad t_2 = 170,\quad t_3 = 210,\quad t_4 = 250,\quad t_5 = 600.$$

Вычислить оценку интенсивности отказов.

Имеем план $[10, Б, T=1000]$ и $d=5$. Для него получаем

$$\lambda_{N,Б,T} = \frac{5}{(120+170+210+250+600)+(10-5)\cdot 1000} = 7{,}874\cdot 10^{-4}\ \text{ч}^{-1}.$$

Задача 54. *Найти оценку интенсивности отказов в условиях задачи* 53, *если испытания приборов были завершены после получения* 5*-го отказа.*

Находим

$$\lambda_{N,Б,r} = \frac{5-1}{(120+170+210+250+600)+(10-5)\cdot 600} = 9{,}195\cdot 10^{-4}\ \text{ч}^{-1}.$$

Таблица 35

Значения коэффициентов b_i (нижняя строка) и ε_i (верхняя строка) [119]

i	k														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0,7968	0,6386	0,5296	0,4514	0,3931	0,3478	0,3121	0,2827	0,2583	0,2381	0,2207	0,2057	0,1924	0,1808	0,1710
	0,6275	0,5232	0,4477	0,3907	0,3463	0,3108	0,2819	0,2579	0,2375	0,2201	0,2053	0,1923	0,1807	0,1703	0,1613
2		0,9266	0,8300	0,7419	0,6670	0,6042	0,5513	0,5066	0,5513	0,5066	0,4063	0,3810	0,3585	0,3384	0,3208
		0,1790	0,2266	0,2361	0,2320	0,2228	0,2119	0,2009	0,2119	0,2009	0,1711	0,1627	0,1551	0,1479	0,1412
3			0,9655	0,9067	0,8434	0,7828	0,7277	0,6782	0,6340	0,5946	0,5596	0,5284	0,5001	0,4745	0,4515
			0,0775	0,1195	0,1402	0,1492	0,1519	0,1511	0,1483	0,1446	0,1402	0,1356	0,1312	0,1268	0,1224
4				0,9810	0,9434	0,8978	0,8047	0,7613	0,7613	0,7212	0,6841	0,6502	0,6191	0,5905	0,5643
				0,0409	0,0709	0,0902	0,1082	0,1116	0,1116	0,1127	0,1124	0,1111	0,1093	0,1073	0,1051
5					0,9885	0,9631	0,9297	0,8928	0,8551	0,8181	0,7827	0,7491	0,7175	0,6880	0,6605
					0,0243	0,0456	0,0615	0,0725	0,0799	0,0847	0,0875	0,0891	0,0896	0,0895	0,0889
6						0,9925	0,9746	0,9496	0,9205	0,8896	0,8583	0,8274	0,7974	0,7686	0,7413
						0,0156	0,0311	0,0438	0,0536	0,0607	0,0659	0,0694	0,0718	0,0733	0,0741
7							0,9948	0,9818	0,0626	0,9394	0,9140	0,8874	0,8606	0,8340	0,8081
							0,0107	0,0222	0,0324	0,0407	0,0472	0,0522	0,0560	0,0587	0,0607
8								0,9963	0,9865	0,9715	0,9528	0,9317	0,9091	0,8858	0,8624
								0,0076	0,0164	0,0246	0,0316	0,0374	0,0421	0,0458	0,0486
9									0,9973	0,9897	0,9778	0,9625	0,9448	0,9255	0,9053
									0,0056	0,0124	0,0191	0,0250	0,0301	0,0344	0,0380
10										0,9979	0,9920	0,9824	0,9697	0,9548	0,9383
										0,0042	0,0097	0,0151	0,0202	0,0247	0,0285
11											0,9984	0,9936	0,9858	0,9752	0,9625
											0,0033	0,0077	0,0122	0,0165	0,0204
12												0,9987	0,9949	0,9883	0,9794
												0,0026	0,0062	0,0100	0,0139
13													0,9990	0,9958	0,9903
													0,0021	0,0051	0,0083
14														0,9991	0,9965
														0,0017	0,0042
15															0,9993
															0,0014

Задача 55. *Получена выборка значений экспоненциально распределенных величин* $x_\imath$: 10,1; 10,6; 11,2; 12,6; 13,4; 14,8; 15,9; 17,1; 19,1. *Необходимо найти уточненную двухстадийную оценку параметра* $\nu = 1/\lambda$ *(см. раздел 2.2.1.2).*

Пусть предполагаемым значением ν является $\nu_0 = 15$. Выбираем уровень достоверности $\alpha = 0{,}95$. Находим по табл. 55 или с помощью апппроксимаций (см. 1.1.8) при $f = 2 \cdot 10 = 20$: $\chi^2\left(\dfrac{1-0{,}95}{2}\right) = \chi^2(0{,}025) = 9{,}59$ и $\chi^2(0{,}975) = 34{,}2$.

Далее вычисляем

$$\bar{x} = \frac{1}{10}\sum_{\imath=1}^{n} x_\imath = 13{,}66 \quad \text{и} \quad \frac{2n\bar{x}}{\nu_0} = \frac{2013{,}66}{15} = 18{,}2133.$$

Убеждаемся, что $9{,}59 < \dfrac{2n\bar{x}}{\nu_0} = 18{,}2133 < 34{,}2$. Так как неравенство выполняется, вычисляем уточненные оценки:

$$\hat{\nu}_1 = \frac{2\cdot 10\cdot 13{,}66}{15\cdot 34{,}2}\cdot 13{,}66 + 15\cdot\left(1 - \frac{2\cdot 10\cdot 13{,}66}{15\cdot 34{,}2}\right) = 14{,}2864;$$

$$\hat{\nu}_2 = \left(\frac{2\cdot 10\cdot 13{,}66}{15\cdot 34{,}2}\right)^2\cdot 13{,}66 + 15\cdot\left(1 - \frac{4\cdot 10^2\cdot 13{,}66^2}{15^2\cdot 34{,}2^2}\right) = 14{,}61769.$$

Задача 56. *Используя данные задачи 55, найти оценки параметра экспоненциального распределения с помощью порядковых статистик.*

Оптимальная линейная оценка (2.2.1.3.2)

Рассматриваем вариант отсутствия цензурирования, т. е. когда $r_1 = r_2 = 0$. В этом случае вычисляем:

$$m = a_1 = \frac{1}{n-1+1} = \frac{1}{n}; \quad l = a_1^2 = \frac{1}{n^2}; \quad c = \left(\frac{1}{n}\right)^2\cdot n^2 + n - 1 = n.$$

Следовательно,

$$\nu_0 = \frac{1}{n}\cdot\left[\left(\frac{n^2}{n} - n\right)\cdot x_1 + \sum_{\imath-1}^{n} x_\imath\right] = \frac{1}{n}\cdot\sum_{\imath-1}^{n} x_\imath = 13{,}66,$$

т. е. оценка совпадает с обычной оценкой максимального правдоподобия.

Предположим, что $r_1 = 2$ первых и $r_2 = 1$ последних членов выборки цензурированы. В этом случае имеем

$$m = \sum_{i=1}^{3} a_\imath; \quad l = \sum_{\imath=1}^{3} a_i^2; \quad a_1 = \frac{1}{n} = \frac{1}{10}; \quad a_2 = \frac{1}{n-1} = \frac{1}{9}; \quad a_3 = \frac{1}{n-2} = \frac{1}{8}.$$

Тогда

$$m = \frac{1}{10} + \frac{1}{9} + \frac{1}{8} = 0{,}33611; \quad l = \frac{1}{100} + \frac{1}{81} + \frac{1}{64} = 0{,}03797;$$

$$c = \frac{(0{,}33611)^2}{0{,}03797} + 10 - 2 - 1 - 1 = 8{,}97524.$$

Предположим теперь, что цензурированы (исключены из выборки) центральные ее члены — известны только $s = 3$ первых члена и $r = 4$ наибольших члена. Для вычисления оценки в этом случае найдем

$$\beta = s + r - 1 + \frac{\left(\sum\limits_{\imath=s+1}^{n-r+1} a_\imath\right)^2}{\sum\limits_{i=s+1}^{n-r+1} a_\imath^2} = 6 + \frac{\left(\sum\limits_{\imath=4}^{7} a_\imath\right)^2}{\sum\limits_{i=4}^{7} a_i^2}.$$

Последовательно вычисляя, получаем

$$a_4 = \frac{1}{n-4+1} = \frac{1}{7}; \quad a_5 = \frac{1}{n-5+1} = \frac{1}{6}; \quad a_6 = \frac{1}{5}; \quad a_7 = \frac{1}{4}; \quad \sum_{i=4}^{7} a_i = 0{,}759523;$$

$$\left(\sum_{i=4}^{7} a_i\right)^2 = 0{,}576876; \quad \sum_{i=4}^{7} a_i^2 = 0{,}1506859; \quad \beta = 6 + \frac{0{,}576876}{0{,}1506859} = 9{,}82833343$$

и окончательно

$$\nu_0^* = \frac{1}{0{,}82833343} \times$$

$$\times \left[\sum_{i=1}^{2} x_i + \left(10 - 3 + 1 + \sum_{i=4}^{7} a_i\right) \cdot x_3 + \left(\frac{0{,}75952}{0{,}1506859} - 4 + 1\right) \cdot x_7 + \sum_{i=1}^{3} x_{7+i}\right] = 20{,}46177.$$

Оценка по одной порядковой статистике (2.2.1.3.2)

Предположим, что оценка производится по 7-й порядковой статистике x_7. Имеем

$$\sum_{i=1}^{7} a_i = \sum_{i=1}^{7} \frac{1}{n-i+1} = \frac{1}{10} + \frac{1}{9} + \ldots + \frac{1}{4} = 1{,}09563492; \quad \nu_7 = \frac{14{,}8}{1{,}09563492} = 13{,}508149.$$

Оценка Эпштейна (2.2.1.3.3)

Предположим, что в выборке цензурированы 2 наименьших наблюдения ($s - 1 = 2$; $s = 3$). Тогда получаем

$$\hat{\nu} = \frac{1}{10} \cdot \left[\left(\frac{3}{\sum\limits_{i=1}^{3} a_i} + 3 - 10\right) \cdot x_3 + \sum_{i=4}^{10} x_i\right] = 0{,}1 \times$$

$$\times \left[\left(\frac{3}{\frac{1}{10} + \frac{1}{9} + \frac{1}{8}} - 7\right) \cdot 11{,}2 + 11{,}8 + 12{,}6 + 13{,}4 + 14{,}8 + 15{,}9 + 17{,}1 + 19{,}1\right] = 12{,}626694.$$

Предположим, что в выборке цензурировано $s - 1 = 2$ наименьших и $r - 1 = 2$ наибольших наблюдений. Тогда оценка будет равна

$$\hat{\nu} = \frac{1}{n-2} \cdot \left[\left(\frac{3}{\sum\limits_{i=1}^{3} a_i} + 3 - 10\right) \cdot x_3 + \sum_{i=4}^{7} x_i + 3 \cdot x_8\right] = 15{,}120867.$$

Оценка Огавы (2.2.1.3.4)

Будем искать оценку по четырем, оптимальным образом выбранным, порядковым статистикам. Из табл. 35 для $k = 4$ находим:

$$\varepsilon_1 = 0{,}4514; \quad \varepsilon_2 = 0{,}7419; \quad \varepsilon_3 = 0{,}9067; \quad \varepsilon_4 = 0{,}9810;$$
$$b_1 = 0{,}3907; \quad b_2 = 0{,}2361; \quad b_3 = 0{,}1195; \quad b_4 = 0,0409.$$

Тогда для оценки отбираем порядковые статистики с номерами

$$[n \cdot 0{,}4514] + 1 = 5; \quad [n \cdot 0{,}7419] + 1 = 7; \quad [n \cdot 0{,}9067] + 1 = 9; \quad [n \cdot 0{,}9810] + 1 = 10.$$

Вычисляем оценку

$$\nu^0 = 0{,}3907 \cdot x_5 + 0{,}2361 \cdot x_7 + 0{,}1195 \cdot x_9 + 0{,}0409 \cdot x_{10} = \\ = 0{,}3907 \cdot 12{,}6 + 0{,}2361 \cdot 14{,}8 + 0{,}1195 \cdot 17{,}1 + 0{,}0409 \cdot 19{,}1 = 11{,}2417.$$

Если бы мы хотели произвести оценку по двум оптимально выбранным порядковым статистикам, то ими должны быть x_7 и x_{10}, для которых $b_1 = 0{,}5232$ и $b_2 = 0{,}1790$.

Тогда оценка равна $\nu^0 = 0{,}5232 \cdot 14{,}8 + 0{,}1790 \cdot 19{,}1 = 11{,}16226$.

2.2.2. Интервальные оценки

Интервальная оценка параметра ν (средняя наработка между отказами) при доверительной вероятности α подсчитывается по формулам

$$\nu_n^{\text{н}} = \frac{2\sum_{i=1}^{n} x_i}{\chi^2_{\gamma'}}; \quad \nu_n^{\text{в}} = \frac{2\sum_{i=1}^{n} x_i}{\chi^2_{\gamma''}},$$

где χ^2_γ — γ-квантиль распределения хи-квадрат с $f = 2n$ степенями свободы; $\gamma' = \dfrac{1+\alpha}{2}$, $\gamma'' = \dfrac{1-\alpha}{2}$ для двусторонней оценки и $\gamma' = \alpha$, $\gamma'' = 1-\alpha$ для односторонней оценки.

На практике интервальные оценки записываются в форме

$$\nu_n^{\text{н}} = \frac{2n\bar{x}}{\chi^2_{\gamma'}}; \quad \nu_n^{\text{в}} = \frac{2n\bar{x}}{\chi^2_{\gamma''}}, \quad \text{где} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Обычно используются табулированные коэффициенты оценок

$$\nu_n^{\text{н}} = \frac{2n\bar{x}}{\chi^2_{\gamma'}}; \quad \nu_n^{\text{в}} = \frac{2n\bar{x}}{\chi^2_{\gamma''}}.$$

Значения коэффициентов $k_{\text{н}}$ и $k_{\text{в}}$ для двусторонней интервальной оценки при $\alpha = 0{,}90$ и $\alpha = 0{,}95$ приведены в табл. 36, заимствованной из [16].

Таблица 36

Значения коэффициентов $k_{\text{н}}$ и $k_{\text{в}}$ [16]

n	$\alpha = 0{,}90$		$\alpha = 0{,}95$		n	$\alpha = 0{,}90$		$\alpha = 0{,}95$	
	$k_{\text{н}}$	$k_{\text{в}}$	$k_{\text{н}}$	$k_{\text{в}}$		$k_{\text{н}}$	$k_{\text{в}}$	$k_{\text{н}}$	$k_{\text{в}}$
1	0,333	19,20	0,270	28,60	16	0,690	1,59	0,645	1,75
2	0,422	5,62	0,360	9,20	17	0,700	1,57	0,655	1,71
3	0,476	3,68	0,420	4,80	18	0,710	1,54	0,660	1,69
4	0,515	2,92	0,455	3,70	19	0,715	1,52	0,665	1,66
5	0,546	2,54	0,480	3,00	20	0,719	1,51	0,675	1,64
6	0,568	2,30	0,515	2,73	25	0,740	1,44	0,700	1,55
7	0,592	2,13	0,535	2,50	30	0,756	1,39	0,720	1,48
8	0,610	2,01	0,555	2,32	40	0,787	1,31	0,750	1,40
9	0,625	1,92	0,575	2,19	50	0,806	1,28	0,770	1,35
10	0,637	1,84	0,585	2,09	70	0,830	1,23	0,800	1,28
11	0,650	1,78	0,598	2,00	100	0,852	1,19	0,830	1,23
12	0,660	1,73	0,610	1,93	200	0,890	1,13	0,870	1,16
13	0,662	1,69	0,620	1,88	300	0,910	1,10	0,895	1,12
14	0,675	1,65	0,630	1,82	500	0,930	1,08	0,915	1,09
15	0,685	1,62	0,640	1,79					

По аналогии с точечными оценками интенсивности отказов ($\lambda = 1/\nu$) для различных планов испытаний на надежность (см. раздел 2.2.1.1), приведем формулы для интервальных оценок.

$$\text{план } [N, B, T]: \quad \lambda_n^{\text{н}} = \frac{c'_\alpha\,(d-1)}{NT}; \quad \lambda_n^{\text{в}} = \frac{c''_\alpha\,(d)}{NT},$$

где c'_α, c''_α — коэффициенты, выражающиеся через квантили распределения Пуассона.

Таблица 37

Значения коэффициентов $c'_\alpha(d)$ и $c''_\alpha(d)$ для $\alpha = 0{,}90$ [140]

d	c'	c''	d	c'	c''	d	c'	c''
0	0,05129	2,99573	6	3,28532	11,84240	12	7,68958	19,44260
1	0,35536	4,74386	7	3,98082	13,14810	13	8,46394	20,66860
2	0,81769	6,29579	8	4,69523	14,43460	14	9,24633	21,88650
3	1,36632	7,75366	9	5,42541	15,70520	15	10,03590	23,09710
4	1,87015	9,15352	10	6,16901	16,99220	16	10,83210	24,30120
5	2,61301	10,51300	11	6,92421	18,20750			

Значения коэффициентов $c'_\alpha(d)$ и $c''_\alpha(d)$ для доверительной вероятности $\alpha = 0{,}90$ в зависимости от числа отказов d приведены в табл. 37, заимствованной из [140].

$$\text{план } [N, B, r]: \quad \lambda_n^{\text{н}} = \frac{c'_\alpha(r-1)}{Nt_r}; \quad \lambda_n^{\text{в}} = \frac{c''_\alpha(r-1)}{Nt_r}.$$

Значения коэффициентов $c'_\alpha(r-1)$ и $c''_\alpha(r-1)$ для $\alpha = 0{,}90$ находятся по табл. 37 при $d = r - 1$.

$$\text{план } [N, Б, r]: \quad \lambda_n^{\text{н}} = \frac{c'_\alpha(r-1)}{\sum\limits_{i=1}^{r} t_i + (N-r)t_r}; \quad \lambda_n^{\text{в}} = \frac{c''_\alpha(r-1)}{\sum\limits_{i=1}^{r} t_i + (N-r)t_r}.$$

$$\text{план } [N, Б, T]: \quad \lambda_n^{\text{н}} = \frac{b'_\alpha(d)}{T}; \quad \lambda_n^{\text{в}} = \frac{b''_\alpha(d)}{T},$$

где $b'_\alpha(d)$ и $b''_\alpha(d)$ — коэффициенты оценки, приведенные в табл. 38 для $\alpha = 0{,}95$ и различных N.

Для плана $[N, B, (r, T)]$ интервальные оценки λ аналогичны оценкам плана $[N, B, T]$, если $t_r > T$, и плана $[N, B, r]$, если $t_r < T$.

$$\text{план } [N, Б, (r, T)]: \quad \lambda_n^{\text{н}} = \frac{b'_\alpha(d)}{T}; \quad \lambda_n^{\text{в}} = \frac{b''_\alpha(d)}{T}, \quad t_r > T,$$

где $b'_\alpha(d)$ и $b''_\alpha(d)$ — коэффициенты, тождественные коэффициентам для плана $[N, Б, T]$.

При $\dfrac{r}{N} < 0{,}1$ планы $[N, B, (r, T)]$ и $[N, Б, (r, T)]$, $[N, B, r]$ и $[N, Б, r]$ практически совпадают [140]. При $\dfrac{d(T)}{T} < 0{,}1$ практически совпадают планы $[N, Б, T]$ и $[N, B, T]$. В указанных условиях можно использовать одинаковые оценки для таких планов.

Для оценок, основанных на порядковых статистиках, укажем один результат — величина $2r\hat{\nu}$ имеет распределение χ^2 с $f = 2r$ степенями свободы. Если в выборке объема n известны r младших наблюдений, а $(n - r)$ старших наблюдений цензу-

Таблица 38

Значения коэффициентов $b'_\alpha(d)$ (нижняя строка) **и $b''_\alpha(d)$** (верхняя строка) **для $\alpha = 0{,}95$** [140]

d	N					
	50	60	80	100	150	200
0	0,0733 0,0000	0,0615 0,0000	0,0461 0,0000	0,0369 0,0000	0,0246 0,0000	0,0184 0,0000
1	0,1126 0,0005	0,0936 0,0004	0,0701 0,0003	0,0560 0,0003	0,0373 0,0002	0,0279 0,0001
2	0,1475 0,0049	0,1225 0,0041	0,0915 0,0031	0,0730 0,0024	0,0485 0,0016	0,0363 0,0012
3	0,1809 0,0126	0,1499 0,0105	0,1117 0,0078	0,0890 0,0063	0,0590 0,0041	0,0422 0,0031
4	0,2136 0,0225	0,1767 0,0186	0,1314 0,0139	0,1045 0,0111	0,0692 0,0073	0,0517 0,0055
5	0,2461 0,0338	0,2032 0,0280	0,1507 0,0208	0,1197 0,0166	0,0791 0,0110	0,0591 0,0082
6	0,2785 0,0464	0,2295 0,0383	0,1697 0,0284	0,1347 0,0226	0,0889 0,0149	0,0663 0,0111
7	0,3111 0,0599	0,2558 0,0494	0,1887 0,0366	0,1496 0,0290	0,0985 0,0191	0,0734 0,0143
8	0,3441 0,0744	0,2823 0,0612	0,2077 0,0452	0,1643 0,0358	0,1080 0,0236	0,0804 0,0176
9	0,3774 0,0896	0,3089 0,0736	0,2267 0,0542	0,1791 0,0429	0,1175 0,0282	0,0874 0,0210
10	0,4130 0,1057	0,3358 0,0866	0,2458 0,0636	0,1938 0,0502	0,1269 0,0330	0,0943 0,0245
11	0,4457 0,1225	0,3630 0,1001	0,2649 0,0733	0,2085 0,0578	0,1361 0,0379	0,1012 0,0282
12	0,4808 0,1399	0,3905 0,1141	0,2842 0,0834	0,2234 0,0657	0,1457 0,0429	0,1081 0,0319
13	0,5166 0,1582	0,4185 0,1286	0,3036 0,0937	0,2383 0,0737	0,1551 0,0481	0,1149 0,0357
14	0,5532 0,1771	0,4469 0,1437	0,3232 0,1044	0,2536 0,0820	0,1644 0,0534	0,1218 0,0396
15	0,5907 0,1968	0,4758 0,1592	0,3429 0,1153	0,2683 0,0904	0,1738 0,0587	0,1286 0,0435
16	0,6292 0,2172	0,5052 0,1752	0,3629 0,1265	0,2834 0,0991	0,1832 0,0642	0,1354 0,0475
17	0,6687 0,1383	0,5352 0,1918	0,3831 0,1380	0,2986 0,1079	0,1927 0,0698	0,1423 0,0516
18	0,7094 0,2603	0,5658 0,2088	0,4036 0,1498	0,3140 0,1169	0,2021 0,0755	0,1491 0,0557
19	0,7513 0,2830	0,5970 0,2264	0,4243 0,1619	0,3295 0,1261	0,2116 0,0812	0,1559 0,0599

рированы, то оптимальная линейная оценка (см. раздел 2.2.1.3.1) имеет вид

$$\hat{\nu} = \frac{1}{r}\left[\sum_{i=1}^{r} x_i + (n-r)\,x_r\right],$$

а интервальные оценки равны $\hat{\nu}_n^{\text{н}} = k_{\text{н}} \cdot \hat{\nu}$; $\nu_n^{\text{в}} = k_{\text{в}} \cdot \hat{\nu}$, где $k_{\text{н}}$ и $k_{\text{в}}$ — коэффициенты, приведенные в табл. 36 (здесь вместо n в таблицу следует входить со значением r).

Задача 57. *Имеются результаты наблюдений над экспоненциально распределенной величиной* x_i: $12, 13, 16, 17, 21, 24, 29, 31, 42, 45, 54$ $(n = 11)$. *Необходимо найти двусторонний доверительный интервал для параметра* ν *при доверительной вероятности* $\alpha = 0{,}95$.

Имеем $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = 27{,}63636$. В табл. 36 для $n = 11$ и $\alpha = 0{,}95$ находим $k_{\text{н}} = 0{,}598$ и $k_{\text{в}} = 2{,}00$. Тогда

$$\nu_n^{\text{н}} = 0{,}598 \cdot 27{,}63636 = 16{,}526; \quad \nu_n^{\text{в}} = 2{,}00 \cdot 27{,}63636 = 55{,}2727.$$

Следовательно, 95%-й доверительный интервал для параметра ν равен

$$16{,}526 \leqslant \nu \leqslant 55{,}2727.$$

Задача 58. *Партия электронных приборов объемом* $N = 100$ *шт. была испытана на надежность с заменой отказавших приборов в течение* 200 *ч. При этом наблюдалось* $d = 5$ *отказов. Необходимо найти двустороннюю* 90%-*ю оценку для интенсивности отказов приборов.*

Имеем план $[100, B, T = 200]$. По таблице 37 находим

$$c'(d-1) = c'(4) = 1{,}97015 \quad \text{и} \quad c''(d) = c''(5) = 10{,}5130.$$

Окончательно имеем

$$\lambda_n^{\text{н}} = \frac{1{,}97015}{100 \cdot 200} = 9{,}8507 \cdot 10^{-5}; \quad \lambda_n^{\text{в}} = \frac{10{,}51303}{100 \cdot 200} = 5{,}2565 \cdot 10^{-4}.$$

Следовательно, 90%-й доверительный интервал для интенсивности отказов равен $9{,}8507 \cdot 10^{-5} \leqslant \lambda \leqslant 5{,}2565 \cdot 10^{-4}$ ч$^{-1}$.

Задача 59. *Испытаны на надежность* 100 *приборов с заменой отказавших приборов. При этом наблюдались* 5 *отказов и момент наступления последнего отказа равен* 212 *ч. Найти* 90%-*й доверительный интервал для интенсивности отказов приборов.*

Имеем план $[100, B, r = 5]$ при $t_r = 212$. Из табл. 37 получаем

$$c'(r-1) = c'(4) = 1{,}97015 \quad \text{и} \quad c''(r-1) = c''(4) = 9{,}15352.$$

Тогда

$$\lambda_n^{\text{н}} = \frac{1{,}97015}{100 \cdot 212} = 9{,}29316 \cdot 10^{-5}; \quad \lambda_n^{\text{в}} = \frac{9{,}15352}{100 \cdot 212} = 4{,}13698 \cdot 10^{-4}.$$

Следовательно, $9{,}29316 \cdot 10^{-5} \leqslant \lambda \leqslant 4{,}13698 \cdot 10^{-4}$.

Задача 60. *Решить задачу* 59 *при условии, что испытания проводились без замены отказавших приборов и моменты наступления отказов были (ч):* $t_1 = 50$, $t_2 = 80$, $t_3 = 110$, $t_4 = 190$ *и* $t_5 = 212$.

Имеем из табл. 37 $c'(r-1) = c'(4) = 1{,}97015$ и $c''(r-1) = c''(4) = 9{,}15352$.
Тогда

$$\sum_{i=1}^{5} t_i = 50 + 80 + 110 + 190 + 212 = 642;$$

$$\lambda_n^{\text{н}} = \frac{1{,}97015}{642 + (100-5) \cdot 212} = 9{,}4800 \cdot 10^{-5}; \quad \lambda_n^{\text{в}} = \frac{9{,}15352}{642 + (100-5) \cdot 212} = 4{,}4045 \cdot 10^{-4};$$

$$9{,}4800 \leqslant \lambda \leqslant 4{,}4045 \cdot 10^{-4}.$$

Задача 61. *Решить задачу 59 при условии, что испытания проводились без замены отказавших приборов.*

Находим для $d = 5$ и $N = 100$ из табл. 38: $b'(5) = 0{,}0166$ и $b''(5) = 0{,}1197$. Тогда

$$\lambda_n^{\text{н}} = \frac{0{,}0166}{200} = 8{,}3 \cdot 10^{-5}; \quad \lambda_n^{\text{в}} = \frac{0{,}1197}{200} = 5{,}985 \cdot 10^{-4};$$

$$8{,}3 \cdot 10^{-5} \leqslant \lambda \leqslant 5{,}985 \cdot 10^{-4}.$$

Задача 62. *Были проведены испытания на безотказность десяти приборов. В результате были получены* 7 *отказов в моменты времени* t_i: 10, 11, 13, 16, 18, 21, 29. *Необходимо найти* 95%*-й доверительный интервал для наработки на отказ.*

Имеем $r = 7$, $n = 10$ и $\hat{\nu} = \frac{1}{7} \cdot \left[\sum_{i=}^{r} x_i + (10 - 7) \cdot 29\right] = 26{,}71428$.
Далее из табл. 36 для $n = r = 7$ имеем $k_{\text{н}} = 0{,}592$ и $k_{\text{в}} = 2{,}13$.
Тогда

$$\nu_n^{\text{н}} = 0{,}592 \cdot 26{,}71428 = 15{,}8148; \quad \nu_n^{\text{в}} = 2{,}13 \cdot 26{,}71428 = 56{,}9014;$$

$$15{,}8148 \leqslant \nu \leqslant 56{,}9014.$$

2.3. Оценка параметров распределения Вейбулла

Напомним основные формы распределения Вейбулла. Закон распределения случайной величины записывается либо в двухпараметрической форме

$$F(x) = 1 - \exp\left\{-\left(\frac{x}{\alpha}\right)^{\beta}\right\},$$

либо в трехпараметрической

$$F(x) = 1 - \exp\left\{-\left(\frac{x-\mu}{\alpha}\right)^{\beta}\right\}.$$

Оценке подлежат либо два параметра: α — параметр масштаба и β — параметр формы, либо три, включая μ — параметр сдвига.

Известно, что случайная величина $y = \ln x$ имеет распределение наименьших значений с функцией

$$F(y) = 1 - \exp\left\{-\exp\left(\frac{y-u}{b}\right)\right\}.$$

Оценки параметров $\hat{u}$ и $\hat{b}$ связаны с оценками $\hat{\alpha}$ и $\hat{\beta}$ соотношениями $\hat{\beta} = \dfrac{1}{\hat{b}}$ и $\hat{\alpha} = e^{\hat{u}}$.

Поэтому на практике часто используется следующий прием. Обработкой ряда величин $\ln x_i$ оцениваются параметры $\hat{u}$ и $\hat{b}$, а затем переходят к оценкам $\hat{\alpha}$ и $\hat{\beta}$.

Особенность распределения Вейбулла — чрезвычайно богатое разнообразие форм кривых распределения — обуславливает его широкое распространение в практике, поэтому совершенствование методов оценки его параметров актуально. Обширный обзор методов оценки параметров распределения Вейбулла приведен в [152].

2.3.1. Точечные оценки

2.3.1.1. Оценка максимального правдоподобия

При известном параметре формы β оценка для α имеет вид

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\beta}\right)^{\frac{1}{\beta}}.$$

Легко видеть, что при $\beta = 1$ имеем оценку $\hat{\alpha} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$, т.е. оценка совпадает с оценкой для параметра экспоненциального распределения, что следует из факта перехода распределения Вейбулла в экспоненциальное при $\beta = 1$ (см. раздел 1.1.5).

При неизвестном β совместные оценки максимального правдоподобия параметров α и β являются решением системы уравнений

$$\begin{cases} n\alpha^{\beta} - \sum\limits_{i=1}^{n} x_i^{\beta} = 0, \\ \dfrac{n}{\beta} + \sum\limits_{i=1}^{n} \ln x_i - \dfrac{1}{\alpha^{\beta}} \sum\limits_{i=1}^{n} x_i^{\beta} \ln x_i = 0. \end{cases}$$

В общем случае эта система решается методом последовательных приближений. Интересный метод ускоренного решения приведенной системы уравнений предложен в [153]. Система сводится к

$$\begin{cases} \alpha = \left(\dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^{\beta}\right)^{\frac{1}{\beta}}, \\ \dfrac{1}{\beta} + \dfrac{1}{n}\sum\limits_{i=1}^{n}\ln x_i - \dfrac{\sum\limits_{i=1}^{n} x_i^{\beta}\ln x_i}{\sum\limits_{i=1}^{n} x_i^{\beta}} = 0. \end{cases}$$

Существование и единственность решения этой системы уравнений показаны в [152]. Система решается методом символических операторов [154] с начальным приближением β_0, исходя из зависимости коэффициента вариации v от β. Точная зависимость $v(\beta)$, определяемая формулой

$$v = \left[\frac{\partial\left(1+\dfrac{2}{\beta}\right)}{\partial^2\left(1+\dfrac{1}{\beta}\right)} - 1\right]^{\frac{1}{2}},$$

достаточно просто аппроксимируется соотношением $v = \beta^{-0{,}93}$ или $\beta_0 = v^{-1{,}075}$ [155] (при $1 \leqslant \beta \leqslant 50$ ошибка менее 3%, при $1 \leqslant \beta \leqslant 25$ — не более 0,4%).

В работе [153] предлагается оценка

$$\hat{\beta} = \left\{\frac{s_3}{s_1} - \frac{s_2}{n} + \beta_0\frac{s_3^2 - s_1 s_4}{s_1^2}\left[\left(\frac{s_3}{s_1} - \frac{s_2}{n}\right)\beta_0 - 1\right]\right\}^{-1},$$

где

$$\beta_0 = v^{-1{,}075}; \quad v = \frac{s}{\bar{x}}; \quad s = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{\frac{1}{2}}; \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad s_1 = \sum_{i=1}^{n} x_i^{\beta_0};$$

$$s_2 = \sum_{i=1}^{n}\ln x_i; \quad s_3 = \sum_{i=1}^{n} x_i^{\beta_0}\ln x_i; \quad s_4 = \sum_{i=1}^{n} x_i^{\beta_0}(\ln x_i)^2.$$

По оценке $\hat{\beta}$ вычисляется оценка параметра α: $\hat{\alpha} = \left(\dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}$.

2.3.1.2. Метод моментов

Метод моментов основан на приравнивании эмпирических моментов статистического ряда их теоретическим значениям, являющимся функциями параметров распределения. Зависимость моментов распределения Вейбулла от его параметров очень сложна (включает в себя комбинацию гамма-функций). Поэтому чаще всего пользуются заранее подготовленными таблицами. Одна из них воспроизведена в табл. 39.

Порядок вычисления оценок $\hat{\alpha}$ и $\hat{\beta}$ включает в себя последовательное вычисление $\bar{x}, s, v, \hat{\beta} = v^{-1{,}075}$ и $\hat{\alpha} = \left(\dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}$.

Таблица 39

Зависимость параметра β распределения Вейбулла от коэффициента вариации $v = s/\bar{x}$ [44, 46]

v	β	v	β	v	β	v	β
3,14	0,400	1,80	0,588	0,910	1,10	0,547	1,90
2,93	0,417	1,67	0,625	0,837	1,20	0,523	2,00
2,75	0,435	1,55	0,667	0,775	1,30	0,496	2,10
2,57	0,455	1,43	0,714	0,723	1,40	0,480	2,20
2,40	0,476	1,32	0,769	0,681	1,50	0,461	2,30
2,24	0,500	1,21	0,833	0,640	1,60	0,444	2,40
2,08	0,526	1,10	0,909	0,605	1,70	0,428	2,50
1,94	0,556	1,00	1,000	0,575	1,80	0,365	3,00

Для трехпараметрического распределения оценка параметров α, β и μ методом моментов рассмотрена в [156]. Авторами рассматривается система трех уравнений, связывающих моменты распределения (α_3 — коэффициент асимметрии, s — среднеквадратическое отклонение и $\bar{x}$ — среднее значение) с параметрами α, β и μ этого распределения. Система имеет вид

$$\begin{cases} \alpha_3 = \dfrac{\Gamma\left(1+\dfrac{3}{\beta}\right) - 3\Gamma\left(1+\dfrac{2}{\beta}\right)\Gamma\left(1+\dfrac{1}{\beta}\right) + 2\Gamma^3\left(1+\dfrac{1}{\beta}\right)}{\left[\Gamma\left(1+\dfrac{2}{\beta}\right) - \Gamma^2\left(1+\dfrac{1}{\beta}\right)\right]^{\frac{3}{2}}}, \\ s = \alpha\left\{\Gamma\left(1+\dfrac{2}{\beta}\right) - \Gamma^2\left(1+\dfrac{1}{\beta}\right)\right\}^{\frac{1}{2}}, \\ \bar{x} = \alpha\Gamma\left(1+\dfrac{1}{\beta}\right) + \mu. \end{cases}$$

Используются оценки

$$\hat{\alpha} = \frac{1}{n}\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^3}{s^3}; \quad s = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{\frac{1}{2}}; \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i.$$

Из зависимости $\alpha_3 = f(\beta)$ и можно найти оценку α_3 и вычислить оценку $\hat{\beta}$. Располагая зависимостью $s = f(\alpha, \beta)$ и зная $\hat{\beta}$, находим оценку $\hat{\alpha}$. Располагая значениями $\hat{\alpha}$ и s, находим оценку μ. Значения необходимых для расчета величин α_3, β, b и c приведены в табл. 40. Схема вычислений включает в себя вычисление α_3 по выборочным данным и определение по табл. 40 соответствующей оценки $\hat{\beta}$, а также коэффициентов b и c. Параметры α и μ оцениваются по формулам $\hat{\alpha} = s \cdot b$; $\mu = \bar{x} - sc$.

Однако этот метод применим только при значительном объеме выборки, так как при $n \leqslant 30$ точность выборочной оценки коэффициента вариации v мала, что приводит к большим погрешностям при оценке параметров распределения. Более точную оценку можно получить, воспользовавшись характеристической порядковой статистикой, как методом моментного сравнения. Очевидно, что для $x = \mu$: $F(x) = 1 - \dfrac{1}{e} = 0{,}632$.

Таблица 40

Значения α_3, β, b' и c, используемые при оценке параметров распределения Вейбулла [156]

α_3	β	b'	c	α_3	β	b'	c
−1,0	40,818	32,827	32,381	1,4	1,267	1,355	1,258
−0,9	22,922	18,858	18,417	1,5	1,211	1,284	1,205
−0,8	15,626	13,153	12,717	1,6	1,160	1,219	1,157
−0,7	11,664	10,046	9,617	1,7	1,115	1,158	1,113
−0,6	9,185	8,094	7,672	1,8	1,073	1,102	1,072
−0,5	7,493	6,755	6,341	1,9	1,035	1,049	1,035
−0,4	6,271	5,780	5,375	2,0	1,000	1,000	1,000
−0,3	5,350	5,039	4,464	2,1	0,968	0,954	0,968
−0,2	4,634	4,456	4,073	2,2	0,939	0,911	0,938
−0,1	4,064	3,986	3,616	2,3	0,911	0,871	0,911
0,0	3,602	3,599	3,243	2,4	0,886	0,833	0,884
0,1	3,222	3,274	2,933	2,5	0,863	0,798	0,860
0,2	2,905	2,997	2,673	2,6	0,841	0,765	0,837
0,3	2,637	2,759	2,452	2,7	0,821	0,734	0,816
0,4	2,410	2,551	2,262	2,8	0,803	0,704	0,796
0,5	2,216	2,368	2,097	2,9	0,785	0,677	0,787
0,6	2,048	2,206	1,954	3,0	0,769	0,650	0,759
0,7	1,802	2,060	1,828	3,1	0,753	0,626	0,743
0,8	1,774	1,929	1,717	3,2	0,739	0,603	0,727
0,9	1,662	1,811	1,618	3,3	0,725	0,581	0,712
1,0	1,564	1,703	1,530	3,4	0,712	0,559	0,697
1,1	1,477	1,605	1,452	3,5	0,700	0,540	0,684
1,2	1,399	1,515	1,381	3,6	0,688	0,521	0,671
1,3	1,329	1,432	1,316	3,7	0,677	0,503	0,658

Пусть $x_{0,632n}$ — порядковая статистика и $r' = 0{,}632(n+1)$.
Тогда $x_R = \sqrt{x_r x_{r+1}}$, где $r < r'$. Рассмотрим систему

$$\begin{cases} x_R = \alpha + \mu, \\ \bar{x} = \alpha\Gamma\left(1 + \dfrac{1}{\beta}\right) + \mu, \\ s = \alpha\left\{\Gamma\left(1 + \dfrac{2}{\beta}\right) - \Gamma^2\left(1 + \dfrac{1}{\beta}\right)\right\}^{-\frac{1}{2}}, \end{cases}$$

откуда, приравнивая

$$G = \frac{x_R - \bar{x}}{s} = \frac{1 - \Gamma\left(1 + \dfrac{1}{\beta}\right)}{\left\{\Gamma\left(1 + \dfrac{2}{\beta}\right) - \Gamma^2\left(1 + \dfrac{1}{\beta}\right)\right\}^{-\frac{1}{2}}},$$

находим оценку $\hat{\beta}$, а затем $\hat{\alpha} = sb'$ и $\mu = x_R - \hat{\alpha}$.

Необходимые значения G, β, b и b' приведены в табл. 41, заимствованной из работы [156]. Вычисляя выборочное значение G, находим по табл. 41 соответствующие ему значения β и b', затем вычисляем оценки $\hat{\alpha} = sb'$ и $\hat{\mu} = x_R - \hat{\alpha}$. Такая оценка вдвое точнее, чем рассмотренная ранее.

Таблица 41

Значения G, β и b' для оценки параметров распределения Вейбулла [156]

G	β	b'	G	β	b'	G	β	b'	G	β	b'
0,00	1,000	1,000	0,11	1,311	1,410	0,22	1,821	1,978	0,33	3,005	3,085
0,01	1,024	1,034	0,12	1,347	1,453	0,23	1,887	2,045	0,34	3,207	3,262
0,02	1,049	1,068	0,13	1,384	1,497	0,24	1,958	2,117	0,35	3,445	3,365
0,03	1,074	1,103	0,14	1,422	1,542	0,25	2,034	2,193	0,36	3,728	3,705
0,04	1,101	1,139	0,15	1,463	1,589	0,26	2,117	2,274	0,37	4,072	3,992
0,05	1,127	1,175	0,16	1,506	1,638	0,27	2,208	2,361	0,38	4,502	4,347
0,06	1,155	1,212	0,17	1,551	1,689	0,28	2,307	2,455	0,39	5,057	4,801
0,07	1,184	1,250	0,18	1,599	1,741	0,29	2,417	2,557	0,40	5,808	5,409
0,08	1,214	1,288	0,19	1,649	1,797	0,30	2,539	2,669	0,41	6,895	6,279
0,09	1,245	1,328	0,20	1,703	1,854	0,31	2,675	2,793	0,42	8,636	7,660
0,10	1,278	1,368	0,21	1,760	1,915	0,32	2,829	2,931	0,43	11,966	10,283

2.3.1.3. Метод наименьших квадратов

Метод рассмотрен в [157]. Идея метода заключается в следующем. Если два раза прологарифмировать плотность распределения Вейбулла $f(t) = \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]$, то получим линейную зависимость вида $y = c + bx$. Оценки плотности вероятностей $f(t_i)$ получаются из выборочной гистограммы

$$P_c = f(t_i) = 1 - \frac{1}{N}\left(\sum_{j=1}^{i} m_j - 0{,}5\right),$$

где m_j — частоты разрядов гистограммы; N — объем выборки.

Для рассматриваемого случая имеют место соотношения

$$y = \lg[-\lg f(t_i), \quad x = \lg t; \quad c = -0{,}3622 - b\lg a.$$

Число таких уравнений равно числу разбиений выборочной гистограммы (предположим, в нашем случае оно равно n). Для отыскания параметров c и b методом наименьших квадратов необходимо решить систему уравнений

$$\begin{cases} nc + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ c\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases}$$

Искомые оценки определяются по формулам

$$c = \frac{D_c}{D_0}; \quad b = \frac{D_b}{D_0}; \quad D_0 = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2; \quad D_c = \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i;$$

$$D_b = n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i.$$

Этот метод позволяет находить параметры распределения Вейбулла непосредственно по статистической гистограмме, не затрачивая время на вычисление его параметров по выборке. Отметим, что можно использовать и графические мето-

ды, идея которых — графическая линеаризация функции распределения Вейбулла путем введения логарифмической шкалы аргумента и двойной логарифмической шкалы функции:

$$\ln\ln\frac{1}{F(x_i)}=\beta\ln x_i-\beta\ln\alpha,\quad \text{где}\quad F(x_i)=\frac{1}{n+1}.$$

Угловой коэффициент такой прямой является оценкой $\hat{\beta}$. Однако графический метод требует точных графических построений (особенно для значений $\beta = 0{,}2 \div 1{,}5$). Для устранения таких трудностей в [158] предложен графоаналитический метод, однако и он может использоваться только для грубой оценки параметров распределения.

В [36] предложена еще одна простая, но достаточно эффективная оценка параметров распределения Вейбулла. Как и ранее, имеем оценки коэффициента асимметрии и стандартного отклонения $\alpha_3=\frac{1}{ns^3}\sum_{i=1}^{n}(x_i-\bar{x})^3$; $s=\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{\frac{1}{2}}$.

Оценки основаны на аппроксимации

$$G=\Gamma\left(1+\frac{1}{\beta}\right)\approx 1-0{,}427(\beta-1)\beta^{-1,9}.$$

При $\beta\geqslant 1$ ошибка аппроксимации $\leqslant 0{,}2\%$.

Оценки вычисляются по формулам

$$\hat{\beta}\approx 4{,}8(\alpha_3+1{,}23)^{-1,4};\quad \hat{\beta}=\frac{\delta}{G};\quad \hat{\mu}=\bar{x}-\delta;\quad \delta\approx\left(0{,}5+0{,}784\hat{\beta}-\frac{0{,}305}{\hat{\beta}}\right)s.$$

При $1{,}5\leqslant\beta\leqslant 20$ ошибка аппроксимации $\leqslant 0{,}7\%$ для $\hat{\beta}$ и $< 0{,}2\%$ для δ.

Для двухпараметрического распределения Вейбулла (когда заранее известно, что $\mu = 0$) оценки параметров β и α имеют вид

$$\hat{\beta}\approx\frac{n-1}{n}\left(0{,}465\frac{s}{\bar{x}}+1{,}282\frac{\bar{x}}{s}-0{,}7\right);\quad \hat{\alpha}=\frac{\bar{x}}{G}.$$

Ошибка аппроксимации $\hat{\beta}$ в этом случае $\leqslant 0{,}25\%$ для $\beta\geqslant 1{,}5$.

2.3.1.4. Оценка с помощью квантилей

Обозначим через x_p и x_q соответственно p- и q-квантили распределения Вейбулла (т. е. $F(x_p)=p$ и $F(x_q)=q$). Тогда оценки параметров распределения Вейбулла имеют вид [159]

$$\hat{\beta}=\frac{\ln d_p-\ln d_q}{\ln x_p-lnx_q};\quad \hat{\alpha}=\exp\left\{-\frac{\ln x\ln d_q-\ln x_q\ln d_p}{\ln d_p-\ln d_q}\right\},$$

где $d_p=-\ln(1-p)$ и $d_q=-\ln(1-q)$.

В [159] показано, что наибольшая эффективность оценки $\hat{\alpha}$ достигается при $p=0{,}398$ и $q=0{,}821$ ($d_p=0{,}5074$ и $d_q=1{,}7203$), а $\hat{\beta}$ — при $p=0{,}167$ и $q=0{,}974$ ($d_p=0{,}1827$ и $d_q=3{,}6496$).

Рекомендуется для совместной оценки параметров $\hat{\alpha}$ и $\hat{\beta}$ использовать квантили уровней $p=0{,}2$ и $0{,}95$ (в этом случае эффективность оценок не менее 60% по сравнению с оценками максимального правдоподобия (см. раздел 2.3.1.1) при всех α и β).

Окончательно рекомендуемые оценки имеют вид

$$\hat{\beta} = -2{,}5973 \ln \frac{x_{0,95}}{x_{0,2}}; \quad \hat{\alpha} = \exp(0{,}4224 \ln x_{0,95}).$$

Квантиль $x_{0,95}$ оценивается порядковой статистикой $x_{[0,05n]+1}$, а $x_{0,2}$ — соответственно $x_{[0,8n]+1}$, где $[\ldots]$ — целая часть числа, заключенного в скобках.

2.3.1.5. Оценки, основанные на порядковых статистиках

Такие оценки наиболее эффективны в обычной практике оценки надежности изделий по данным о наработке первых r отказавших приборов из общего числа n испытываемых. В прикладной математической статистике такая задача формулируется как задача оценки параметров распределения вероятностей по цензурированной сверху выборке (при оценке не учитываются $(n-r)$ наибольших по величине членов выборки).

Оценки для цензурированных выборок, основанные на линейной комбинации порядковых статистик (простые линейные оценки), рассмотрены в [160–164]. В [165] рассмотрена задача оценки параметров распределения Вейбулла при прогрессивном цензурировании, когда часть изделий снимается с испытаний после каждого отказа.

Мы рассмотрим наиболее простые для практического применения и достаточно эффективные наилучшие линейные оценки.

Напомним, что если случайная величина x имеет двухпараметрическое распределение Вейбулла с функцией $F(x) = 1 - \exp\left\{-\left(\frac{x}{\alpha}\right)^{\beta}\right\}$, то случайная величина $y = \ln x$ будет иметь распределение наименьших значений с функцией

$$F(y) = 1 - \exp\left\{-\exp\left(\frac{y-u}{b}\right)\right\}.$$

Суть метода заключается в поиске параметров распределения величины $\ln x$ ($\hat{u}$ и $\hat{b}$) в форме

$$\hat{u} = \sum_{i=1}^{r} a_i y_i; \quad \hat{b} = \sum_{i=1}^{r} c_i y_i$$

с последующим переходом к оценке параметров распределения исходной величины x (не следует забывать, что оценка проводится по первым r наблюдениям из выборки объема n).

Оценки $\hat{u}$ и $\hat{b}$ являются смещенными, и для исключения смещения используются поправки, с учетом которых несмещенные оценки равны

$$\hat{u}^* = \hat{u} + \breve{b}^* k_1; \quad \hat{b}^* = \frac{b^*}{1-k_2}.$$

Значения коэффициентов a_i, c_i, k_1 и k_2 приведены в табл. 42 и 43, заимствованных из [95].

2.3.1.6. Оценка параметров распределения Рэлея (частный случай распределения Вейбулла)

Распределение Рэлея (см. раздел 1.1.15) является частным случаем распределения Вейбулла при $\beta = 2$. Значение $\beta = 2$ является граничным между регулярным и нерегулярным случаями распределения Вейбулла и заслуживает отдельного рассмотрения. Достаточно полно это сделано в [166].

Таблица 42

Значения коэффициентов a_i и c_i [95]

n	r	i	a_i	c_i	n	r	i	a_i	c_i
2	2	1	0,110731	−0,421383	7	3	1	−0,272195	−0,315369
		2	0,889269	0,421383			2	−0,184061	−0,281139
3	2	1	−0,166001	−0,452110			3	1,456255	0,596507
		2	1,166001	0,452110	7	4	1	−0,110274	−0,229691
3	3	1	0,081063	−0,278666			2	−0,060226	−0,215611
		2	0,251001	−0,190239			3	0,018671	−0,164168
		3	0,667936	0,468904			4	1,151829	0,609472
4	2	1	−0,346974	−0,465455	7	5	1	−0,030368	−0,176203
		2	1,346974	0,465455			2	0,004333	−0,172398
4	3	1	−0,044975	−0,297651			3	0,052957	−0,141218
		2	0,088057	−0,234054			4	0,117599	−0,082820
		3	0,956918	0,531705			5	0,855480	0,572640
4	4	1	0,064336	−0,203052	7	6	1	0,013524	−0,138436
		2	0,147340	−0,182749			2	0,041588	−0,140342
		3	0,261510	−0,070109			3	0,075499	−0,121821
		4	0,526813	0,455910			4	0,117461	−0,082994
5	2	1	−0,481434	−0,472962			5	0,172092	−0,015394
		2	1,481434	0,472962			6	0,579835	0,498931
5	3	1	−0,137958	−0,306562	7	7	1	0,038743	−0,108323
		2	−0,025510	−0,257087			2	0,064086	−0,113479
		3	1,164680	0,563650			3	0,090785	−0,103569
5	4	1	−0,006983	−0,217766			4	0,120971	−0,078748
		2	0,059652	−0,199351			5	0,157657	−0,032632
		3	0,156664	−0,118927			6	0,207825	0,054727
		4	0,790668	0,536044			7	0,319934	0,382022
5	5	1	0,052975	−0,158131	8	2	1	−0,752513	−0,483616
		2	0,103531	−0,155707			2	1,752513	0,483616
		3	0,163808	−0,111820	8	3	1	−0,323875	−0,317890
		4	0,246092	−0,005600			2	−0,243808	−0,288231
		5	0,433593	0,431259			3	1,566830	0,606120
6	2	1	−0,588298	−0,477782	8	4	1	−0,149973	−0,232805
		2	1,588298	0,477782			2	−0,105015	−0,220324
6	3	1	−0,211474	−0,311847			3	−0,032257	−0,176675
		2	−0,112994	−0,271381			4	1,287245	0,629805
		3	1,324468	0,583229	8	5	1	−0,062656	−0,180231
6	4	1	−0,063569	−0,225141			2	−0,032248	−0,176510
		2	−0,006726	−0,209083			3	0,012767	−0,149566
		3	0,079882	−0,146386			4	0,072446	−0,101642
		4	0,990412	0,580610			5	1,009691	0,607948
6	5	1	0,007521	−0,169920	8	6	1	−0,013509	−0,143834
		2	0,048328	−0,166319			2	0,010292	−0,145006
		3	0,101608	−0,129510			3	0,041357	−0,128393
		4	0,172859	−0,054453			4	0,080475	−0,095696
		5	0,669685	0,520201			5	0,130327	−0,043280
6	6	1	0,048826	−0,128810			6	0,751058	0,556209
		2	0,079377	−0,132102	8	7	1	0,015973	−0,116317
		3	0,117541	−0,111951			2	0,036729	−0,120331
		4	0,163591	−0,064666			3	0,060439	−0,110582
		5	0,226486	0,031796			4	0,088239	−0,088450
		6	0,368179	0,405733			5	0,122062	−0,050995
7	2	1	−0,676874	−0,481140			6	0,165529	0,009700
		2	1,676874	0,481140			7	0,511030	0,476975

Продолжение таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
8	8	1	0,034052	−0,093270	10	2	1	−0,876869	−0,487022
		2	0,053552	−0,098886			2	1,876869	0,487022
		3	0,073452	−0,093994	10	3	1	−0,408602	−0,321265
		4	0,095062	0,079752			2	−0,340443	−0,297858
		5	0,119768	−0,053918			3	1,749045	0,619124
		6	0,149934	−0,010179	10	4	1	−0,214930	−0,236817
		7	0,191236	0,069325			2	−0,177223	−0,226688
		8	0,282943	0,360675			3	−0,113820	−0,193159
9	2	1	−0,818444	−0,485517			4	1,505973	0,656663
		2	1,818444	0,485517	10	5	1	0,115524	−0,185169
9	3	1	−0,368833	−0,319786			2	−0,090868	−0,181821
		2	−0,296280	−0,293621			3	−0,511341	−0,160698
		3	1,664113	0,613407			4	0,000925	−0,125311
9	4	1	−0,184461	−0,235080			5	1,256809	0,652997
		2	−0,143505	−0,223891	10	6	1	−0,058017	−0,149985
		3	−0,075815	−0,185970			2	−0,039595	−0,150451
		4	1,403781	0,644941			3	−0,012513	−0,136941
9	5	1	−0,907260	−0,183061			4	0,022314	−0,112224
		2	−0,063541	−0,179515			5	0,065750	−0,075721
		3	−0,021495	−0,155825			6	1,022062	0,625321
		4	0,034259	−0,115133	10	7	1	−0,022198	−0,124170
		5	1,141604	0,633534			2	−0,006900	−0,126894
9	6	1	−0,037118	−0,147411			3	0,013224	−0,118392
		2	−0,016377	−0,148150			4	0,037994	−0,100924
		3	0,012499	−0,133219			5	0,068153	−0,073988
		4	0,049305	−0,105060			6	0,105164	−0,035501
		5	0,095614	−0,062073			7	0,804572	0,579868
		6	0,896078	0,595913	10	8	1	0,001179	−0,104082
9	7	1	−0,004220	−0,120988			2	0,014889	−0,108163
		2	0,013386	−0,124245			3	0,030998	−0,103119
		3	0,035068	−0,115091			4	0,049734	−0,090835
		4	0,061198	−0,095508			5	0,071745	−0,070902
		5	0,093013	−0,064162			6	0,096114	−0,041560
		6	0,153879	−0,038125			7	0,130649	−0,000799
		7	0,647676	0,558119			8	0,602692	0,517864
9	8	1	0,016797	−0,100011	10	9	1	0,016841	−0,087358
		2	0,032919	−0,104750			2	0,029807	−0,092405
		3	0,050582	−0,099608			3	0,043570	−0,089839
		4	0,070497	−0,086226			4	0,058640	−0,081428
		5	0,093635	−0,063541			5	0,075576	−0,066855
		6	0,121560	−0,028346			6	0,095169	−0,044670
		7	0,157175	0,026525			7	0,118707	−0,011816
		8	0,456836	0,455956			8	0,148575	0,038159
9	9	1	0,030338	−0,081777			9	0,413116	0,436394
		2	0,048720	−0,087308	10	10	1	0,027331	−0,072734
		3	0,061368	−0,085084			2	0,040034	−0,077971
		4	0,077742	−0,076470			3	0,052496	−0,077242
		5	0,095769	−0,060667			4	0,065408	−0,071876
		6	0,116517	−0,035136			5	0,079263	−0,061652
		7	0,141932	0,006001			6	0,094638	−0,045420
		8	0,176764	0,078828			7	0,112414	−0,020698
		9	0,253697	0,341614			8	0,134239	0,017927

Продолжение таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
10	10	9	0,164178	0,085070	11	10	7	0,094625	−0,030120
		10	0,230001	0,324597			8	0,114811	0,000537
11	2	1	−0,929310	−0,488243			9	0,140333	0,044638
		2	1,920310	0,488243			10	0,377130	0,418384
11	3	1	−0,444245	−0,322452	11	11	1	0,024850	−0,065444
		2	−0,380642	−0,301277			2	0,035456	−0,070318
		3	1,824887	0,623620			3	0,045727	−0,070456
11	4	1	−0,242206	−0,238188			4	0,056215	−0,067076
		2	−0,207204	−0,228941			5	0,067261	−0,060207
		3	−0,147490	−0,198888			6	0,079220	−0,049300
		4	1,596900	0,666017			7	0,092560	−0,033156
11	5	1	−0,137718	−0,186803			8	0,108034	−0,009427
		2	−0,115110	−0,183651			9	0,127068	0,026879
		3	−0,077762	−0,164597			10	0,153197	0,089148
		4	−0,028411	−0,133278			11	0,210412	0,309357
		5	1,359000	0,668329	12	2	1	−0,976872	−0,489254
11	6	1	−0,076739	−0,151936			2	1,976872	0,489254
		2	−0,060142	−0,152221	12	3	1	−0,476530	−0,323426
		3	−0,034581	−0,139907			2	−0,416836	−0,304093
		4	−0,001490	−0,117886			3	1,893367	0,617519
		5	0,039518	−0,086131	12	4	1	−0,266888	−0,239300
		6	1,133434	0,648081			2	−0,234180	−0,230796
11	7	1	−0,038349	−0,126507			3	−0,177681	−0,203562
		2	−0,024842	−0,128838			4	1,678749	0,673657
		3	−0,005964	−0,120951	12	5	1	−0,157792	−0,188109
		4	0,017632	−0,105219			2	−0,136684	−0,185012
		5	0,046354	−0,081602			3	−0,101445	−0,167790
		6	0,081182	−0,048929			4	−0,054640	−0,136930
		7	0,923987	0,612047			5	1,450761	0,680734
11	8	1	−0,012943	−0,106922	12	6	1	−0,093679	−0,053471
		2	−0,001050	−0,110498			2	−0,078561	−0,153632
		3	0,013869	−0,105662			3	−0,054320	−0,142329
		4	0,031661	−0,094405			4	−0,022769	−0,122474
		5	0,052723	−0,076693			5	0,016136	−0,094355
		6	0,077815	−0,051525			6	1,233193	0,666261
		7	0,108161	−0,016860	12	7	1	−0,052987	−0,128308
		8	0,729765	0,562564			2	−0,040893	−0,130339
11	9	1	0,004425	−0,091115			3	−0,023072	−0,123007
		2	0,015498	−0,095437			4	−0,000515	−0,108712
		3	0,028023	−0,092780			5	0,026930	−0,087681
		4	0,042178	−0,084833			6	0,059918	−0,059256
		5	0,058340	−0,071581			7	1,030620	0,637304
		6	0,077093	−0,052182	12	8	1	−0,025785	−0,109045
		7	0,099349	−0,024880			2	−0,015312	−0,112224
		8	0,126592	0,013606			3	−0,001353	−0,107627
		9	0,548502	0,499202			4	0,015634	−0,097276
11	10	1	0,016502	−0,077717			5	0,035853	−0,081361
		2	0,027205	−0,082449			6	0,059835	−0,059315
		3	0,038291	−0,081388			7	0,088444	−0,029900
		4	0,050160	−0,075977			8	0,842684	0,596748
		5	0,063170	−0,066222	12	9	1	−0,006944	−0,093658
		6	0,077772	−0,051429			2	0,002669	−0,097540

Продолжение таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
12	9	3	0,014239	−0,094893	13	5	4	−0,078337	−0,144971
		4	0,027669	−0,087448			5	1,533976	0,690983
		5	0,043189	−0,075374	13	6	1	−0,109187	−0,172412
		6	0,061225	−0,058180			2	−0,092014	−0,168148
		7	0,082441	−0,034802			3	−0,076615	−0,144215
		8	0,107856	−0,003342			4	−0,041997	−0,101104
		9	0,667655	0,545234			5	−0,004940	−0,001512
12	10	1	0,006411	−0,080881			6	1,323488	0,711124
		2	0,015598	−0,085171	13	7	1	−0,066358	−0,129743
		3	0,025675	−0,083952			2	−0,055414	−0,131538
		4	0,036799	−0,078714			3	−0,038503	−0,124701
		5	0,049211	−0,067610			4	−0,016879	−0,111609
		6	0,063256	−0,056237			5	0,009416	−0,092649
		7	0,079438	−0,037675			6	0,040810	−0,067475
		8	0,098522	−0,012272			7	1,126930	0,657714
		9	0,121752	0,022956	13	8	1	−0,037540	−0,110704
		10	0,503338	0,481555			2	−0,028206	−0,113563
12	11	1	0,015982	−0,069798			3	−0,015049	−0,109206
		2	0,024997	−0,074285			4	−0,001123	−0,099644
		3	0,034156	−0,074131			5	0,020686	−0,085204
		4	0,043790	−0,070617			6	0,043677	−0,065581
		5	0,054149	−0,063891			7	0,070830	−0,039995
		6	0,065515	−0,053621			8	0,944372	0,623896
		7	0,078264	−0,039034	13	9	1	−0,017389	−0,095590
		8	0,092958	−0,018715			2	−0,008934	−0,099109
		9	0,110521	0,009948			3	0,001863	−0,096521
		10	0,132666	0,052280			4	0,014684	−0,089554
		11	0,347003	0,401864			5	0,029637	−0,078490
12	12	1	0,022771	−0,059449			6	0,047027	−0,063068
		2	0,031776	−0,063952			7	0,067346	−0,046607
		3	0,040408	−0,064601			8	0,091328	−0,015928
		4	0,049122	−0,062489			9	0,774437	0,580865
		5	0,058175	−0,037754	13	10	1	−0,002927	−0,083170
		6	0,067800	−0,050137			2	0,005067	−0,087085
		7	0,078281	−0,039010			3	0,014356	−0,085792
		8	0,090017	−0,023199			4	0,024891	−0,080789
		9	0,103664	−0,000505			5	0,036816	−0,072325
		10	0,120475	0,033696			6	0,050389	−0,060181
		11	0,143566	0,091751			7	0,065995	−0,043768
		12	0,193947	0,295648			8	0,084201	−0,022048
13	2	1	−1,020378	−0,490105			9	0,105863	0,006715
		2	2,020378	0,490105			10	0,615348	0,528441
13	3	1	−0,506031	−0,324239	13	11	1	0,007628	−0,072617
		2	−0,449735	−0,306454			2	0,015408	−0,076746
		3	1,955765	0,630694			3	0,023732	−0,076418
13	4	1	−0,289420	−0,240219			4	0,032743	−0,072938
		2	−0,258687	−0,232349			5	0,042611	−0,066531
		3	−0,205024	−0,207450			6	0,053556	−0,057014
		4	1,753131	0,630018			7	0,065876	−0,043886
13	5	1	−0,176109	−0,189177			8	0,080005	−0,026244
		2	−0,156637	−0,186381			9	0,096594	−0,002552
		3	−0,122893	−0,170454			10	0,116703	0,029910

Продолжение таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
13	11	11	0,465143	0,465037	14	7	7	1,214708	0,674581
13	12	1	0,015382	−0,063288	14	8	1	−0,048365	−0,112041
		2	0,023100	−0,067492			2	−0,039964	−0,114637
		3	0,030818	−0,067892			3	−0,027495	−0,110509
		4	0,038824	−0,065622			4	−0,011849	−0,101635
		5	0,047302	−0,060887			5	0,006905	−0,088422
		6	0,056444	−0,053540			6	0,029002	−0,088422
		7	0,066482	−0,043158			7	0,054897	−0,048074
		8	0,077739	−0,028970			8	1,036868	0,640520
		9	0,090699	−0,009644	14	9	1	−0,027030	−0,097117
		10	0,106166	0,017233			2	−0,019516	−0,100334
		11	0,125627	0,056547			3	−0,009363	−0,097827
		12	0,321416	0,386713			4	0,002928	−0,091298
13	13	1	0,021005	−0,054436			5	0,017368	−0,081103
		2	0,028757	−0,058585			6	0,034165	−0,067124
		3	0,036127	−0,059535			7	0,053685	−0,048921
		4	0,043501	−0,058259			8	0,076476	−0,025720
		5	0,051078	−0,054942			9	0,871287	0,609445
		6	0,059028	−0,049472	14	10	1	−0,011580	−0,084931
		7	0,067533	−0,041504			2	−0,004548	−0,088528
		8	0,076831	−0,030980			3	−0,004548	−0,087207
		9	0,087274	−0,015037			4	0,014100	−0,082451
		10	0,099441	0,006644			5	0,025647	−0,074573
		11	0,114446	0,038943			6	0,038794	−0,063473
		12	0,135068	0,093324			7	0,053879	−0,048768
		13	0,179913	0,283257			8	0,071335	−0,029776
14	2	1	−1,060461	−0,490831			9	0,091783	−0,005398
		2	2,060461	0,490831			10	0,716445	0,565101
14	3	1	−0,533185	−0,324929	14	11	1	−0,000170	−0,074686
		2	−0,479874	−0,308462			2	0,006622	−0,078499
		3	2,013059	0,633391			3	0,014283	−0,078064
14	4	1	−0,310144	−0,240982			4	0,022800	−0,074680
		2	−0,281132	−0,233670			5	0,032273	−0,068624
		3	−0,229990	−0,210735			6	0,042866	−0,059816
		4	1,821266	0,685397			7	0,054817	−0,047926
14	5	1	−0,192947	−0,190068			8	0,068463	−0,032355
		2	−0,174709	−0,187427			9	0,084290	−0,012126
		3	−0,142478	−0,172710			10	0,103025	0,014349
		4	−0,099930	−0,149393			11	0,570731	0,512429
		5	1,610065	0,699598	14	12	1	0,008361	−0,065816
14	6	1	−0,123352	−0,155736			2	0,015058	−0,069728
		2	−0,110490	−0,155747			3	0,022076	−0,099620
		3	−0,088443	−0,146054			4	0,029552	−0,067659
		4	−0,059523	−0,120460			5	0,037615	−0,063070
		5	−0,024111	−0,106556			6	0,046411	−0,056130
		6	1,405919	0,693553			7	0,056132	−0,046558
14	7	1	−0,078656	−0,130915			8	0,067039	−0,033834
		2	−0,068666	−0,132521			9	0,079506	−0,017101
		3	−0,052554	−0,126123			10	0,094096	0,005064
		4	−0,031776	−0,114051			11	0,111723	0,035156
		5	−0,006522	−0,096788			12	0,432431	0,449638
		6	0,023467	−0,074184	14	13	1	0,014760	−0,057849

Продолжение таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
14	13	2	0,021453	−0,061764	16	7	7	1,369798	0,700883
		3	0,028054	−0,062506	16	8	1	−0,067719	−0,114069
		4	0,034842	−0,061704			2	−0,060754	−0,116260
		5	0,041933	−0,057693			3	−0,049415	−0,112545
		6	0,049474	−0,052317			4	−0,034868	−0,104798
		7	0,057619	−0,044707			5	−0,011736	−0,093508
		8	0,066569	−0,034420			6	0,003178	−0,078726
		9	0,076605	−0,020713			7	0,026973	−0,060251
		10	0,088151	−0,002338			8	1,199963	0,680158
		11	0,101914	0,022943	16	9	1	−0,044303	−0,099396
		12	0,119200	0,059643			2	−0,038218	−0,102138
		13	0,299416	0,372795			3	−0,029094	−0,099811
14	14	1	0,019487	−0,050186			4	−0,017697	−0,094037
		2	0,026238	−0,054008			5	−0,004166	−0,085242
		3	0,032614	−0,055130			6	0,011570	−0,073467
		4	0,038941	−0,054419			7	0,028712	−0,058535
		5	0,045399	−0,052075			8	0,050576	−0,040084
		6	0,052097	−0,048606			9	1,041619	0,652711
		7	0,059168	−0,042197	16	10	1	−0,027135	−0,087496
		8	0,066767	−0,034099			2	−0,021550	−0,090585
		9	0,075102	−0,022315			3	−0,013895	−0,089277
		10	0,084482	−0,008285			4	−0,004646	−0,084992
		11	0,095428	0,012430			5	0,006132	−0,078105
		12	0,108942	0,043015			6	0,018515	−0,068653
		13	0,127523	0,094166			7	0,032675	−0,056482
		14	0,167807	0,272004			8	0,048869	−0,041268
16	2	1	−1,123243	−0,492005			9	0,067459	−0,022503
		2	2,132243	0,492005			10	0,893576	0,619360
16	3	1	−0,581757	−0,326035	16	11	1	−0,014263	−0,077597
		2	−0,533457	−0,311694			2	−0,008950	−0,080895
		3	2,115214	0,637730			3	−0,002286	−0,080349
16	4	1	−0,347172	−0,242220			4	−0,002286	−0,077313
		2	−0,321026	−0,235794			5	0,005469	−0,071820
		3	−0,274186	−0,215984			6	0,014703	−0,064207
		4	1,942384	0,693998			7	0,024296	−0,054237
16	5	1	−0,223015	−0,191470			8	0,035593	−0,041635
		2	−0,206788	−0,189099			9	0,048404	−0,025917
		3	−0,177158	−0,176323			10	0,063020	−0,006432
		4	−0,138048	−0,156390			11	0,079847	0,580293
		5	1,745009	0,713282	16	12	1	−0,004450	−0,069172
16	6	1	−0,148725	−0,157331			2	0,000732	−0,072584
		2	−0,137508	−0,157263			3	0,006721	−0,072615
		3	−0,117232	−0,148785			4	0,013424	−0,070383
		4	−0,090481	−0,134532			5	0,020868	−0,066184
		5	−0,057883	−0,115196			6	0,029314	−0,060054
		6	1,551828	0,713108			7	0,038344	−0,051876
16	7	1	−0,100621	−0,132718			8	0,048668	−0,041398
		2	−0,092121	−0,134040			9	0,060342	−0,028216
		3	−0,077354	−0,128381			10	0,073692	−0,011716
		4	−0,058057	−0,117942			11	0,089173	−0,009035
		5	−0,034624	−0,103296			12	0,623351	0,535164
		6	−0,007020	−0,084506	16	13	1	0,003118	−0,061843

Продолжение таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
16	13	2	0,008356	−0,065297	16	16	12	0,079051	0,002079
		3	0,013789	−0,065770			13	0,088111	0,021044
		4	0,019747	−0,064259			14	0,099315	0,048675
		5	0,026189	−0,061031			15	0,114733	0,094419
		6	0,033196	−0,056120			16	0,147977	0,252333
		7	0,040872	−0,049427	18	2	1	−1,195128	−0,492912
		8	0,049357	−0,040731			2	2,195128	0,492912
		9	0,058836	−0,029675	18	3	1	−0,624252	−0,326884
		10	0,069568	−0,015710			2	−0,580008	−0,314183
		11	0,081920	0,002010			3	2,204260	0,641066
		12	0,096438	0,024833	18	4	1	−0,379529	−0,243153
		13	0,498713	0,483018			2	−0,355679	−0,237429
16	14	1	0,008992	−0,055309			3	−0,312382	−0,219992
		2	0,014141	−0,058750			4	2,047590	0,700574
		3	0,019370	−0,059563	18	5	1	−0,249266	−0,192523
		4	0,024804	−0,058635			2	−0,234618	−0,190376
		5	0,030525	−0,056208			3	−0,207148	−0,179091
		6	0,036615	−0,052317			4	−0,170883	−0,161679
		7	0,043164	−0,046867			5	1,861914	0,723670
		8	0,050284	−0,039699	18	6	1	−0,170868	−0,158516
		9	0,058124	−0,030467			2	−0,160910	−0,158405
		10	0,066884	−0,018695			3	−0,142100	−0,150876
		11	0,076854	−0,003625			4	−0,117175	−0,138383
		12	0,088469	−0,015969			5	−0,086906	−0,121647
		13	0,102433	0,042224			6	1,677960	0,727829
		14	0,379341	0,421953	18	7	1	−0,119793	−0,134044
16	15	1	0,013547	−0,049291			2	−0,112406	−0,135163
		2	0,018743	−0,052670			3	−0,098738	−0,130098
		3	0,023778	−0,053739			4	−0,080698	−0,120904
		4	0,028849	−0,053290			5	−0,058807	−0,108183
		5	0,034060	−0,051538			6	−0,033165	−0,092095
		6	0,039489	−0,048520			7	1,503605	0,720486
		7	0,045218	−0,044164	18	8	1	−0,084626	−0,115541
		8	0,051338	−0,038307			2	−0,078711	−0,117434
		9	0,057965	−0,030678			3	−0,068272	−0,114068
		10	0,065253	−0,020850			4	−0,054656	−0,107202
		11	0,074425	−0,008156			5	−0,038217	−0,097349
		12	0,082818	0,008503			6	−0,019006	−0,084645
		13	0,093994	0,031075			7	0,003084	−0,069301
		14	0,107995	0,063476			8	1,340405	0,705270
		15	0,263528	0,348149	18	9	1	−0,059414	−0,101022
16	16	1	0,017016	−0,043375			2	−0,054359	−0,103411
		2	0,022284	−0,046633			3	−0,046030	−0,101260
		3	0,027208	−0,047890			4	−0,035375	−0,096099
		4	0,032046	−0,047839			5	−0,022631	−0,088374
		5	0,036918	−0,046675			6	−0,007819	−0,078203
		6	0,041887	−0,044432			7	0,009161	−0,065532
		7	0,047042	−0,041053			8	0,028495	−0,050186
		8	0,052455	−0,036402			9	1,187973	0,684087
		9	0,058216	−0,030249	18	10	1	−0,040776	−0,089291
		10	0,064444	−0,022230			2	−0,036223	−0,091997
		11	0,071304	−0,011772			3	−0,029314	−0,090739

Продолжение таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
18	10	4	−0,020701	−0,086863	18	14	10	0,052564	−0,026314
		5	−0,010540	−0,080764			11	0,061904	−0,014497
		6	−0,001172	−0,072544			12	0,072509	0,000034
		7	0,014523	−0,062157			13	0,084730	0,018080
		8	0,029671	−0,049445			14	0,552004	0,507970
		9	0,046841	−0,034147	18	15	1	0,004780	−0,052617
		10	1,045347	0,657947			2	0,008843	−0,055674
18	11	1	−0,026669	−0,079582			3	0,013074	−0,056526
		2	−0,022402	−0,082484			4	0,017522	−0,055995
		3	−0,016466	−0,081896			5	0,022232	−0,054191
		4	−0,009294	−0,079012			6	0,027249	−0,051331
		5	−0,000979	−0,074183			7	0,032630	−0,047281
		6	0,008496	−0,067503			8	0,038443	−0,042029
		7	0,019212	−0,058930			9	0,044772	−0,035403
		8	0,031300	−0,048324			10	0,051728	−0,027176
		9	0,044847	−0,035451			11	0,059460	−0,017018
		10	0,060404	−0,019962			12	0,068169	−0,004442
		11	0,911490	0,627325			13	0,078145	−0,011289
18	12	1	−0,015793	−0,071378			14	0,089813	0,031340
		2	−0,011677	−0,074393			15	0,443142	0,456986
		3	−0,006416	−0,074315	18	16	1	0,009048	−0,047594
		4	−0,000278	−0,072211			2	0,013157	−0,050597
		5	0,006695	−0,068395			3	0,017235	−0,051629
		6	0,014529	−0,062952			4	0,021396	−0,051393
		7	0,023297	−0,055848			5	0,025706	−0,050102
		8	0,033110	−0,046959			6	0,030212	−0,047820
		9	0,044122	−0,036073			7	0,034966	−0,044532
		10	0,056540	−0,022877			8	0,040027	−0,040165
		11	0,070367	−0,006924			9	0,045465	−0,034587
		12	0,785235	0,592326			10	0,051368	−0,027599
18	13	1	−0,007289	−0,064317			11	0,057855	−0,028906
		2	−0,003238	−0,067387			12	0,065087	−0,008069
		3	0,001550	−0,067701			13	0,073294	0,005581
		4	0,006940	0,066218			14	0,082827	0,023119
		5	0,012925	−0,063222			15	0,094248	0,046408
		6	0,019540	−0,058792			16	0,338109	0,397887
		7	0,026851	−0,052898	18	17	1	0,012444	−0,042879
		8	0,034951	−0,045430			2	0,016611	−0,045800
		9	0,043969	−0,036200			3	0,020593	−0,046965
		10	0,054072	−0,024926			4	0,024555	−0,047008
		11	0,065486	−0,011201			5	0,028573	−0,046121
		12	0,078516	0,005561			6	0,032702	−0,044362
		13	0,665728	0,552731			7	0,036990	−0,041722
18	14	1	0,000568	−0,058133			8	0,041487	−0,038137
		2	0,003471	−0,061213			9	0,046252	−0,033494
		3	0,007930	−0,061830			10	0,051355	−0,027618
		4	0,012775	−0,060849			11	0,056889	−0,020248
		5	0,018027	−0,058527			12	0,062980	−0,010994
		6	0,023730	−0,054936			13	0,069810	−0,000742
		7	0,029942	−0,050053			14	0,077655	0,015937
		8	0,036744	−0,043781			15	0,086979	0,036313
		9	0,044239	−0,035952			16	0,098636	0,065331

Продолжение таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
18	17	17	0,235490	0,327023	20	8	7	−0,017755	−0,075681
18	18	1	0,015092	−0,098116			8	1,463585	0,724575
		2	0,019328	−0,038165	20	9	1	−0,072826	−0,102246
		3	0,023258	−0,042221			2	−0,068644	−0,104362
		4	0,027089	−0,042947			3	−0,060858	−0,102371
		5	0,030909	−0,041963			4	−0,050834	−0,097711
		6	0,034773	−0,040676			5	−0,038781	−0,090828
		7	0,038728	−0,038627			6	−0,024779	−0,081874
		8	0,042820	−0,035765			7	−0,008798	−0,070863
		9	0,047095	−0,031992			8	0,009270	−0,057714
		10	0,051612	−0,027160			9	1,316151	0,707969
		11	0,056443	−0,021041	20	10	1	−0,052900	−0,090626
		12	0,061685	−0,013300			2	−0,049115	−0,093031
		13	0,067477	−0,003410			3	−0,042792	−0,091837
		14	0,074032	−0,009488			4	−0,034710	−0,088309
		15	0,081713	0,026940			5	−0,025087	−0,082842
		16	0,091221	0,052132			6	−0,013973	−0,075573
		17	0,109314	0,093529			7	−0,001335	−0,077511
		18	0,132411	0,235693			8	0,012921	−0,055584
20	2	1	−1,251068	−0,493634			9	0,028939	−0,042651
		2	2,251068	0,493634			10	1,178052	0,686964
20	3	1	−0,662014	−0,327555	20	11	1	−0,037716	−0,081036
		2	−0,621129	−0,316157			2	−0,034222	−0,083625
		3	2,283144	0,643712			3	−0,028845	−0,083028
20	4	1	−0,408252	−0,243885			4	−0,022146	−0,076014
		2	−0,386289	−0,238726			5	−0,014276	−0,076014
		3	−0,345972	−0,223155			6	−0,005262	−0,070070
		4	2,140513	0,705766			7	0,004930	−0,062554
20	5	1	−0,272551	−0,193344			8	0,016382	−0,053398
		2	−0,259179	−0,191335			9	0,029216	−0,042476
		3	−0,233536	−0,181278			10	0,043593	−0,029594
		4	−0,199675	−0,165821			11	1,048347	0,662168
		5	1,964941	0,731828	20	12	1	−0,025922	−0,072964
20	6	1	−0,190502	−0,159435			2	−0,022589	−0,075662
		2	−0,181539	−0,159298			3	−0,017879	−0,075522
		3	−0,163969	−0,152528			4	−0,012183	−0,073554
		4	−0,140605	−0,141408			5	−0,005600	−0,070076
		5	−0,112303	−0,126651			6	0,001858	−0,065197
		6	1,788917	0,739321			7	0,010227	−0,058928
20	7	1	−0,136790	−0,135060			8	0,019578	−0,051211
		2	−0,130201	−0,136029			9	0,030012	−0,041931
		3	−0,117518	−0,131448			10	0,041668	−0,030912
		4	−0,100561	−0,123236			11	0,054724	−0,017911
		5	−0,079995	−0,111990			12	0,926107	0,633868
		6	−0,056007	−0,097918	20	13	1	−0,016619	−0,066052
		7	1,621135	0,735681			2	−0,013364	−0,068809
20	8	1	−0,099621	−0,116659			3	−0,009129	−0,069021
		2	−0,094504	−0,118326			4	−0,004170	−0,067601
		3	−0,084808	−0,115235			5	0,001453	−0,064335
		4	−0,071993	−0,109093			6	0,007742	−0,060825
		5	−0,056488	−0,100352			7	0,014732	−0,055581
		6	−0,038416	−0,089210			8	0,022485	−0,049051

Продолжение таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
20	13	9	0,031087	−0,041132	20	17	3	0,012297	−0,049380
		10	0,040654	−0,031666			4	0,015773	−0,049304
		11	0,051333	−0,020429			5	0,019394	−0,048357
		12	0,063321	−0,007116			6	0,023192	−0,046613
		13	0,810474	0,602120			7	0,027201	−0,044083
20	14	1	−0,009191	−0,060043			8	0,031459	−0,040736
		2	−0,005961	−0,062821			9	0,036010	−0,036510
		3	−0,002065	−0,063307			10	0,040910	−0,031298
		4	0,002348	−0,062329			11	0,046228	−0,024954
		5	0,007248	−0,060148			12	0,052055	−0,017265
		6	0,012649	−0,056856			13	0,058509	−0,007934
		7	0,018585	−0,052465			14	0,065756	0,003474
		8	0,025109	−0,046929			15	0,074029	0,017606
		9	0,032297	−0,040154			16	0,089671	0,035486
		10	0,040241	−0,031999			17	0,398968	0,433947
		11	0,049069	−0,022261	20	18	1	0,008847	−0,041706
		12	0,058941	−0,011999			2	0,012215	−0,044331
		13	0,070007	−0,003196			3	0,015502	−0,045422
		14	0,700660	0,566775			4	0,018813	−0,045550
20	15	1	−0,003203	−0,054744			5	0,022197	−0,044896
		2	0,000035	−0,575130			6	0,025690	−0,043529
		3	0,003690	−0,058213			7	0,029324	−0,041460
		4	0,007695	−0,057596			8	0,033136	−0,038666
		5	0,012048	−0,055899			9	0,037162	−0,035086
		6	0,016769	−0,053209			10	0,041450	−0,030632
		7	0,021892	−0,049537			11	0,046055	−0,025168
		8	0,027467	−0,044884			12	0,051050	−0,018506
		9	0,033552	−0,039043			13	0,056532	−0,010374
		10	0,040230	−0,032010			14	0,062634	−0,000381
		11	0,047602	−0,023560			15	0,069547	0,012071
		12	0,055804	−0,013436			16	0,077558	0,027938
		13	0,065012	−0,001280			17	0,087131	0,048871
		14	0,075467	0,013416			18	0,305157	0,376826
		15	0,595940	0,527466	20	19	1	0,011469	−0,037905
20	16	1	0,001656	−0,050002			2	0,014895	−0,040446
		2	0,004926	−0,052742			3	0,018135	−0,041607
		3	0,008410	−0,053608			4	0,021329	−0,041903
		4	0,012111	−0,053287			5	0,024554	−0,041503
		5	0,016046	−0,051997			6	0,027802	−0,040467
		6	0,020247	−0,049816			7	0,031153	−0,038810
		7	0,024742	−0,046757			8	0,034624	−0,036509
		8	0,029575	−0,042785			9	0,038247	−0,033514
		9	0,034801	−0,037825			10	0,042061	−0,029746
		10	0,040482	−0,031763			11	0,046112	−0,025085
		11	0,046707	−0,024342			12	0,050458	−0,019364
		12	0,053585	−0,0l5601			13	0,055176	−0,012340
		13	0,061262	−0,004939			14	0,060372	−0,003660
		14	0,069938	−0,008023			15	0,066198	0,007217
		15	0,079893	0,023984			16	0,072887	0,021167
		16	0,495616	0,483548			17	0,080830	0,039737
20	17	1	0,005617	−0,045695			18	0,090746	0,066024
		2	0,008931	−0,048385			19	0,212971	0,308714

Окончание таблицы 42

n	r	i	a_i	c_i	n	r	i	a_i	c_i
20	20	1	0,013555	−0,034055	20	20	11	0,046357	−0,024632
		2	0,017039	−0,036484			12	0,050215	−0,019814
		3	0,020257	−0,037686			13	0,054354	−0,013860
		4	0,023370	−0,098123			14	0,058856	−0,006460
		5	0,026464	−0,037945			15	0,063842	0,002866
		6	0,029565	−0,037211			16	0,069496	0,014902
		7	0,032711	−0,035932			17	0,076128	0,031052
		8	0,035932	−0,034091			18	0,084346	0,054203
		9	0,039258	−0,031016			19	0,095669	0,092028
		10	0,042720	−0,028527			20	0,119862	0,221415

Таблица 43

Значения коэффициентов k_1 и k_2 [95]

n	r	k_1	k_2	n	r	k_1	k_2
2	2	0,037574	0,415839	10	2	0,902322	0,486871
3	2	0,257509	0,450055		3	0,417951	0,315415
	3	−0,018421	0,256346		4	0,229478	0,229309
4	2	0,413509	0,464388		5	0,120330	0,177275
	3	0,084775	0,281729		6	0,062998	0,142198
	4	−0,028312	0,183862		7	0,027627	0,116706
5	2	0,533791	0,472308		8	0,004749	0,097048
	3	0,166129	0,294192		9	−0,010438	0,081004
	4	0,030763	0,202419		10	−0,020508	0,066792
	5	−0,029135	0,142830	11	2	0,952399	0,488120
6	2	0,631490	0,477340		3	0,452207	0,317159
	3	0,232697	0,301732		4	0,246536	0,231380
	4	0,080351	0,212422		5	0,141299	0,179627
	5	0,008881	0,156905		6	0,080450	0,144834
	6	−0,027716	0,116577		7	0,042460	0,116760
7	2	0,713665	0,480082		8	0,017512	0,100437
	3	0,288854	0,306813		9	0,000584	0,085031
	4	0,122608	0,218847		10	−0,011097	0,060304
	5	0,042126	0,164973		11	−0,019101	0,060304
	6	−0,001300	0,127606	12	2	0,997998	0,489151
	7	−0,025789	0,098365		3	0,483387	0,318593
8	2	0,784533	0,483377		4	0,270268	0,233072
	3	0,337341	0,310476		5	0,160246	0,181531
	4	0,159281	0,223358		6	0,096410	0,146945
	5	0,071292	0,170378		7	0,056079	0,122006
	6	0,022472	0,134224		8	0,029301	0,103043
	7	−0,006413	0,107264		9	0,010873	0,087994
	8	−0,023866	0,025017		10	−0,002107	0,075575
9	2	0,846604	0,485329		11	−0,011349	0,064873
	3	0,379959	0,313246		12	−0,017385	0,054954
	4	0,191609	0,226712	13	2	1,039851	0,490018
	5	0,097153	0,174294		3	0,511988	0,319794
	6	0,043783	0,138801		4	0,292046	0,234480
	7	0,011395	0,112788		5	0,177997	0,183107
	8	−0,009069	0,092365		6	0,111099	0,148677
	9	−0,022094	0,074824		7	0,068047	0,123901

Окончание таблицы 43

n	r	k_1	k_2	n	r	k_1	k_2
13	8	0,040225	0,105124	18	3	0,627488	0,323709
	9	0,020463	0,090302		4	0,308007	0,239034
	10	0,006357	0,078188		5	0,249075	0,188145
	11	−0,003882	0,067951		6	0,170681	0,154141
	12	−0,011361	0,058952		7	0,119838	0,129778
	13	−0,016749	0,050470		8	0,084973	0,111433
14	2	1,078521	0,490757		9	0,060068	0,097092
	3	0,538401	0,320813		10	0,041720	0,085544
	4	0,312161	0,235672		11	0,027876	0,076015
	5	0,194239	0,184433		12	0,017236	0,067989
	6	0,124694	0,150126		13	0,008940	0,061101
	7	0,080302	0,025473		14	0,002403	0,055086
	8	0,050382	0,106830		15	−0,002784	0,049739
	9	0,029469	0,092166		16	−0,006912	0,044879
	10	0,014307	0,080248		17	−0,010185	0,040334
	11	0,003200	0,070275		18	−0,012727	0,035808
	12	−0,005064	0,061682	20	2	1,263654	0,493598
	13	−0,011233	0,053001		3	0,664622	0,324060
	14	−0,015764	0,046657		4	0,408277	0,240184
16	2	1,480153	0,491948		5	0,271937	0,189405
	3	0,585828	0,322450		6	0,189878	0,155492
	4	0,348282	0,237577		7	0,136375	0,131212
	5	0,223426	0,186541		8	0,099482	0,112948
	6	0,149158	0,152413		9	0,072972	0,098688
	7	0,101317	0,127935		10	0,053316	0,087226
	8	0,068748	0,109474		11	0,038380	0,077792
	9	0,045666	0,095010		12	0,026809	0,069871
	10	0,028811	0,083327		13	0,017704	0,063107
	11	0,016221	0,073645		14	0,010451	0,057241
	12	0,006658	0,065435		15	0,004619	0,052208
	13	−0,000690	0,058318		16	−0,000103	0,047471
	14	−0,006373	0,051997		17	−0,003943	0,043295
	15	−0,010762	0,046198		18	−0,007067	0,039437
	16	−0,014090	0,040524		19	−0,009596	0,035771
18	2	1,209127	0,492867		20	−0,011599	0,032070

Запишем функцию распределения Рэлея в двухпараметрической форме:

$$F(x;\mu) = 1 - \exp\left\{-\left(\frac{x-\mu}{a}\right)^2\right\}.$$

Оценки параметров μ и a методом моментов имеют вид

$$\hat{a}_M = \frac{2s}{\sqrt{4-\pi}}; \quad \hat{\mu}_M = \bar{x} - \frac{s\sqrt{\pi}}{\sqrt{4-\pi}}, \quad \text{где} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Их дисперсии равны:

$$D(\hat{a}_M) = \frac{4}{4-\pi}k_1; \quad D(\hat{\mu}_M) = k_2 + \frac{4}{4-\pi}k_1 - 2k_3\frac{\sqrt{\pi}}{\sqrt{4-\pi}},$$

$$\text{где} \quad k_1 = \frac{a(4+2\pi-\pi^2)}{4n(4-\pi)}; \quad k_2 = \frac{a^2(4-\pi)}{4n}; \quad k_3 = \frac{a^2(\pi-3)\sqrt{\pi}}{4n\sqrt{4-\pi}}.$$

Оптимальные квантильные оценки имеют вид [166]

$$\hat{a}_K = \frac{x_{p_1} - x_{p_2}}{\alpha_1 - \alpha_2}; \quad \hat{\mu}_K = \frac{\alpha_1 x_{p_2} - \alpha_2 x_{p_1}}{\alpha_1 - \alpha_2},$$

где x_{p_i} — выборочные квантили уровня p_i; $\alpha_i = -\ln(1 - p_i)$, $i = 1, 2$; $p_1 = 0{,}93$ и $p_2 = 0{,}07$.

Дисперсии таких оценок равны

$$D(\hat{a}_K) = 0{,}7681\frac{a^2}{n}; \quad D(\hat{\mu}_K) = 1{,}7506\frac{a^2}{n}.$$

Эффективность моментных и квантильных оценок по сравнению с оценками максимального правдоподобия равны $\approx 0{,}45$ для $\hat{a}_M$ и $\approx 0{,}33$ для $\hat{a}_K$.

2.3.2. Интервальные оценки

2.3.2.1. Оценка α при известном β

Интервальные оценки имеют вид

$$\hat{\alpha}_n^{\text{н}} = \left(\frac{2\sum_{i=1}^{n} x_i^{\beta}}{\chi^2_{\gamma'}}\right)^{\frac{1}{\beta}}; \quad \hat{\alpha}_n^{\text{в}} = \left(\frac{2\sum_{i=1}^{n} x_i^{\beta}}{\chi^2_{\gamma''}}\right)^{\frac{1}{\beta}},$$

где χ^2_γ — γ-квантиль распределения хи-квадрат с $f = 2n$ (n — объем выборки) степенями свободы; $\gamma' = \dfrac{1+\alpha}{2}$, $\gamma'' = \dfrac{1-\alpha}{2}$ для двусторонних оценок и $\gamma' = \alpha$, $\gamma'' = 1 - \alpha$ для односторонних оценок (α — доверительная вероятность, которую не следует путать с α — параметром распределения).

Для нахождения этих оценок можно использовать результаты, полученные ранее для оценки параметра экспоненциального распределения (см. раздел 2.2). По аналогии искомые оценки можно записать в форме

$$\hat{\alpha}_n^{\text{н}} = (k_{\text{н}})^{\frac{1}{\beta}}\tilde{x}^{\frac{1}{\beta}}; \quad \hat{\alpha}_n^{\text{в}} = (k_{\text{в}})^{\frac{1}{\beta}}\tilde{x}^{\frac{1}{\beta}},$$

где $k_{\text{н}}$ и $k_{\text{в}}$ — коэффициенты оценок, табулированные в табл. 36;

$$\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\beta}.$$

В [36] предложены достаточно точные и простые аппроксимации для нахождения интервальных оценок параметров распределения Вейбулла при доверительной вероятности $\alpha = 0{,}95$:

$$\frac{\hat{\beta}}{k_n} \leqslant \beta \leqslant \hat{\beta}k_n; \quad \hat{\alpha}q_1 \leqslant \alpha \leqslant \hat{\alpha}q_2,$$

где $\hat{\beta}$, $\hat{\alpha}$ — точечные оценки соответственно параметров β и α,

$$k_n = 1 + \frac{2{,}05}{(n-3)^{0{,}55}} \text{ (при } n \geqslant 6 \text{ ошибка не более } 0{,}2\%);$$

$$q_1 = \left[1 - \frac{1{,}659}{(n+3)^{0{,}4675}}\right]^{\frac{1}{\hat{\beta}}}; \quad q_2 = \left[1 + \frac{3{,}01}{(n-3{,}15)^{0{,}5623}}\right]^{\frac{1}{\hat{\beta}}} \text{ (ошибка меньше } 1\%).$$

2.3.2.2. Совместная интервальная оценка параметров α и β

Манн и Фертиг [95, 162, 167] предложили оценки для α и β, основанные на их точечных линейных оценках по цензурированным сверху выборкам (см. раздел 2.3.1.5). Оценки имеют вид

$$\hat{\beta}_n^{\text{н}} = c'_\alpha \hat{\beta}; \quad \hat{\beta}_n^{\text{в}} = c''_\alpha \hat{\beta},$$

где $\hat{\beta}$ — точечная линейная оценка параметра β, а c'_α и c''_α — коэффициенты оценки; α — доверительная вероятность.

Значения коэффициентов c'_α и c''_α для различных доверительных вероятностей α приведены в табл. 44.

Интервальные оценки для параметра α находятся по формулам

$$\hat{\alpha}_n^{\text{н}} = \exp(\hat{u} - \hat{b}d'_\alpha); \quad \hat{\alpha}_n^{\text{в}} = \exp(\hat{u} - \hat{b}d''_\alpha),$$

где $\hat{u}$, $\hat{b}$ — оценки параметров распределения наименьших значений (напомним, что $\hat{u} = \sum\limits_{i=1}^{n} a_i \ln x_i$, $\hat{b} = \sum\limits_{i=1}^{n} c_i \ln x_i$ с коэффициентами a_i и c_i из табл. 42); d'_α, d''_α — коэффициенты оценок, α — доверительная вероятность. Значения коэффициентов d'_α и d''_α приведены в табл. 45.

Для распределения Вейбулла представляют интерес оценки, связанные с его широким применением при исследовании надежности технических компонентов и систем. Например, в практике часто возникает задача, которую можно сформулировать следующим образом: найти значение наработки t_ε компонента (или системы), при котором вероятность безотказной работы компонента (или системы) равна заданной величине ε.

Точечная оценка подсчитывается по формуле $t_\varepsilon = -\hat{\alpha}(\ln\varepsilon)^{\frac{1}{\hat{\beta}}}$, где $\hat{\alpha}$, $\hat{\beta}$ — точечные оценки параметров распределения.

Интервальная оценка находится по формуле $\exp\big[\hat{u} - \hat{b}k'(\varepsilon,\alpha)\big] \leqslant t_\varepsilon \leqslant \exp\big[\hat{u} - \hat{b}k''(\varepsilon,\alpha)\big]$, где $\hat{u} = \sum\limits_{i=1}^{r} a_i \ln t_i$, $\hat{b} = \sum\limits_{i=1}^{r} c_i \ln t_i$ — линейные оценки параметров u и b с коэффициентами a_i и c_i из табл. 42; $k'(\varepsilon,\alpha)$ и $k''(\varepsilon,\alpha)$ — коэффициенты оценки и α — доверительная вероятность. Значения коэффициентов $k'(\varepsilon,\alpha)$ и $k''(\varepsilon,\alpha)$ приведены в табл. 46.

Учитывая чрезвычайно широкое применение распределения Вейбулла и его частного случая — экспоненциального распределения — при планировании и оценке результатов испытаний на надежность, в заключение попытаемся ответить еще на один интересный в практическом отношении вопрос: зачем ставить на испытания $n > r$ изделий, если оценка параметров производится по результатам первых r отказов? Ответ прост (хотя и не тривиален) — для экономии времени. Не вдаваясь в математические тонкости (о них желающий может узнать из [95]), приведем известные результаты расчетов показателя $\eta = \dfrac{t(r,n)}{t(r,r)}$, где $t(r,n)$ — ожидаемая продолжительность испытаний n изделий до появления r-го отказа и $t(r,r)$ — ожидаемая продолжительность испытаний, когда на испытания ставятся r изделий и испытания проводятся до отказа всех изделий.

Значения η для разных r и n приведены в табл. 47 для случая экспоненциального распределения (таблица заимствована из [95]). Табл. 47 применима и при оценках, связанных с распределением Вейбулла (при известном заранее параметре β). В этом случае следует найденную по табл. 47 величину η заменить величиной $\eta' = \eta^{\frac{1}{\beta}}$.

Таблица 44

Значения коэффициентов c'_α и c''_α [95]

n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$		n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$	
		c'	c''	c'	c''			c'	c''	c'	c''
3	3	0,17	1,56	0,11	1,86	12	7	0,41	1,47	0,34	1,66
4	3	0,15	1,56	0,10	1,90		8	0,46	1,45	0,38	1,61
	4	0,28	1,53	0,20	1,77		9	0,50	1,43	0,42	1,58
5	3	0,14	1,59	0,09	1,93		10	0,53	1,40	0,45	1,55
	4	0,26	1,55	0,18	1,82		11	0,56	1,37	0,49	1,51
	5	0,36	1,50	0,28	1,70		12	0,60	1,35	0,53	1,46
6	3	0,14	1,59	0,09	1,92	13	3	0,13	1,58	0,08	1,95
	4	0,25	1,55	0,18	1,84		4	0,22	1,57	0,15	1,86
	5	0,33	1,51	0,25	1,73		5	0,30	1,55	0,22	1,79
	6	0,41	1,46	0,33	1,64		6	0,36	1,51	0,28	1,72
7	3	0,14	1,56	0,08	1,92		7	0,40	1,48	0,33	1,67
	4	0,24	1,54	0,17	1,82		8	0,45	1,42	0,37	1,62
	5	0,32	1,52	0,25	1,75		9	0,49	1,40	0,42	1,58
	6	0,39	1,48	0,32	1,67		10	0,52	1,38	0,44	1,55
	7	0,46	1,48	0,38	1,60		11	0,55	1,36	0,48	1,51
8	3	0,13	1,58	0,03	1,95		12	0,58	1,33	0,51	1,47
	4	0,23	1,55	0,16	1,83		13	0,61	1,58	0,54	1,44
	5	0,31	1,52	0,23	1,76	14	3	0,13	1,57	0,08	1,94
	6	0,38	1,49	0,30	1,69		4	0,22	1,54	0,16	1,86
	7	0,44	1,45	0,36	1,62		5	0,30	1,51	0,22	1,77
	8	0,50	1,41	0,42	1,56		6	0,35	1,48	0,28	1,71
9	3	0,13	1,58	0,08	1,92		7	0,40	1,45	0,33	1,67
	4	0,23	1,55	0,16	1,84		8	0,45	1,42	0,38	1,63
	5	0,31	1,52	0,23	1,76		9	0,49	1,40	0,41	1,59
	6	0,38	1,48	0,30	1,70		10	0,52	1,38	0,45	1,56
	7	0,43	1,46	0,35	1,65		11	0,55	1,36	0,48	1,52
	8	0,48	1,42	0,40	1,59		12	0,57	1,34	0,50	1,49
	9	0,53	1,39	0,45	1,53		13	0,60	1,32	0,53	1,46
10	3	0,13	1,59	0,08	1,92		14	0,63	1,57	0,57	1,43
	4	0,23	1,57	0,16	1,86	15	3	0,13	1,56	0,08	1,92
	5	0,30	1,53	0,23	1,77		4	0,22	1,53	0,16	1,85
	6	0,37	1,49	0,29	1,71		5	0,29	1,50	0,22	1,79
	7	0,42	1,46	0,36	1,66		6	0,35	1,48	0,28	1,71
	8	0,47	1,43	0,39	1,60		7	0,41	1,45	0,33	1,67
	9	0,51	1,40	0,43	1,55		8	0,45	1,43	0,37	1,63
	10	0,55	1,38	0,48	1,51		9	0,49	1,41	0,41	1,59
11	3	0,13	1,60	0,08	1,97		10	0,52	1,39	0,42	1,56
	4	0,22	1,58	0,15	1,87		11	0,54	1,37	0,48	1,54
	5	0,30	1,54	0,22	1,82		12	0,57	1,35	0,50	1,50
	6	0,36	1,52	0,28	1,73		13	0,59	1,33	0,52	1,47
	7	0,41	1,48	0,33	1,67		14	0,62	1,32	0,55	1,45
	8	0,46	1,45	0,38	1,62		15	0,64	1,58	0,58	1,42
	9	0,50	1,42	0,42	1,58	16	3	0,13	1,56	0,08	1,94
	10	0,54	1,38	0,46	1,53		4	0,22	1,54	0,15	1,86
	11	0,57	1,36	0,50	1,49		5	0,29	1,51	0,22	1,78
12	3	0,13	1,56	0,08	1,87		6	0,35	1,47	0,27	1,74
	4	0,22	1,55	0,16	1,82		7	0,40	1,45	0,31	1,69
	5	0,30	1,53	0,23	1,78		8	0,44	1,43	0,36	1,64
	6	0,36	1,49	0,29	1,72		9	0,48	1,41	0,40	1,60

Окончание таблицы 44

n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$		n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$	
		c'	c''	c'	c''			c'	c''	c'	c''
16	10	0,51	1,39	0,43	1,57	18	17	0,65	1,30	0,59	1,39
	11	0,53	1,38	0,46	1,54		18	0,67	1,28	0,61	1,37
	12	0,56	1,36	0,49	1,50	20	3	0,12	1,60	0,07	1,97
	13	0,59	1,34	0,51	1,48		4	0,22	1,57	0,15	1,89
	14	0,61	1,34	0,54	1,46		5	0,29	1,55	0,22	1,81
	15	0,63	1,32	0,56	1,43		6	0,35	1,52	0,27	1,75
	16	0,65	1,29	0,59	1,40		7	0,40	1,49	0,32	1,70
18	3	0,12	1,59	0,07	1,93		8	0,43	1,46	0,35	1,65
	4	0,22	1,58	0,15	1,87		9	0,47	1,44	0,39	1,61
	5	0,29	1,54	0,22	1,79		10	0,50	1,42	0,43	1,58
	6	0,35	1,51	0,27	1,73		11	0,53	1,40	0,45	1,54
	7	0,40	1,48	0,33	1,68		12	0,55	1,38	0,48	1,52
	8	0,44	1,46	0,37	1,63		13	0,57	1,36	0,50	1,50
	9	0,47	1,43	0,40	1,59		14	0,60	1,35	0,52	1,48
	10	0,51	1,41	0,43	1,56		15	0,61	1,34	0,54	1,46
	11	0,53	1,39	0,46	1,54		16	0,63	1,33	0,56	1,44
	12	0,56	1,38	0,49	1,52		17	0,65	1,31	0,58	1,42
	13	0,58	1,36	0,51	1,48		18	0,66	1,30	0,60	1,40
	14	0,60	1,35	0,54	1,46		19	0,68	1,28	0,62	1,37
	15	0,62	1,33	0,56	1,44		20	0,70	1,27	0,64	1,36
	16	0,64	1,31	0,57	1,42						

Таблица 45

Значения коэффициентов d'_α и d''_α [95]

n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$		n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$	
		d'	d''	d'	d''			d'	d''	d'	d''
3	3	2,12	−2,54	3,39	−4,47	9	4	0,79	−3,78	1,00	−6,31
4	3	1,55	−3,85	2,43	−6,92		5	0,76	−2,10	0,98	−3,19
	4	1,49	−1,50	2,15	−2,37		6	0,76	−1,38	0,99	−2,01
5	3	1,20	−5,22	1,76	−9,35		7	0,76	−0,99	0,99	−1,43
	4	1,22	−1,94	1,74	−3,13		8	0,76	−0,76	0,99	−1,08
	5	1,20	−1,08	1,64	−1,63		9	0,76	−0,64	0,98	−0,87
6	3	1,02	−6,12	1,39	−10,54	10	3	0,87	−9,98	1,07	−17,45
	4	1,03	−2,39	1,42	−3,69		4	0,77	−4,17	0,96	−6,54
	5	1,04	−1,36	1,41	−2,05		5	0,72	−2,37	0,93	−3,56
	6	1,04	−0,91	1,39	−1,29		6	0,71	−1,51	0,92	−2,21
7	3	0,90	−7,39	1,20	−13,00		7	0,70	−1,08	0,93	−1,56
	4	0,89	−2,95	1,20	−4,67		8	0,71	−0,86	0,93	−1,20
	5	0,89	−1,59	1,21	−2,48		9	0,71	−0,70	0,93	−0,97
	6	0,90	−1,04	1,20	−1,54		10	0,71	−0,60	0,92	−0,80
	7	0,90	−0,79	1,18	−1,09	11	3	0,87	−10,68	1,07	−18,52
8	3	0,88	−8,15	1,72	−14,26		4	0,75	−4,57	0,92	−7,26
	4	0,83	−3,30	1,07	−5,34		5	0,69	−2,58	0,88	−4,00
	5	0,82	−1,86	1,07	−2,78		6	0,66	−1,67	0,85	−2,45
	6	0,82	−1,20	1,08	−1,81		7	0,65	−1,21	0,86	−1,70
	7	0,82	−0,88	1,08	−1,28		8	0,65	−0,92	0,86	−1,30
	8	0,82	−0,70	1,07	−0,97		9	0,65	−0,76	0,86	−1,06
9	3	0,86	−9,12	1,06	−15,68		10	0,65	−0,63	0,86	−0,87

Окончание таблицы 45

n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$		n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$	
		d'	d''	d'	d''			d'	d''	d'	d''
11	11	0,65	−0,55	0,85	−0,75	16	6	0,63	−2,34	0,77	−3,34
12	3	0,88	−11,23	1,10	−19,08		7	0,59	−1,68	0,73	−2,42
	4	0,75	−4,81	0,92	−7,44		8	0,56	−1,30	0,70	−1,81
	5	0,68	−2,72	0,84	−4,17		9	0,54	−1,05	0,68	−1,44
	6	0,64	−1,83	0,81	−2,63		10	0,53	−0,86	0,67	−1,18
	7	0,62	−1,32	0,80	−1,91		11	0,52	−0,72	0,66	−1,00
	8	0,62	−1,00	0,79	−1,41		12	0,52	−0,64	0,66	−0,87
	9	0,62	−0,80	0,80	−1,15		13	0,52	−0,56	0,66	−0,76
	10	0,62	−0,67	0,80	−0,91		14	0,52	−0,51	0,66	−0,68
	11	0,62	−0,58	0,80	−0,78		15	0,52	−0,46	0,66	−0,61
	12	0,62	−0,53	0,79	−0,69		16	0,52	−0,44	0,66	−0,56
13	3	0,88	−11,66	1,09	−19,77	18	3	0,97	−14,29	1,21	−25,90
	4	0,76	−5,21	0,93	−8,22		4	0,83	−6,23	1,00	−9,67
	5	0,68	−2,95	0,84	−4,44		5	0,72	−3,74	0,87	−5,55
	6	0,64	−1,94	0,79	−2,86		6	0,65	−2,56	0,78	−3,67
	7	0,61	−1,40	0,77	−2,04		7	0,60	−1,87	0,72	−2,64
	8	0,59	−1,06	0,75	−1,52		8	0,55	−1,42	0,68	−2,02
	9	0,58	−0,86	0,74	−1,18		9	0,52	−1,16	0,65	−1,62
	10	0,58	−0,72	0,74	−1,00		10	0,50	−0,95	0,63	−1,33
	11	0,58	−0,63	0,75	−0,85		11	0,49	−0,81	0,62	−1,13
	12	0,59	−0,56	0,75	−0,74		12	0,48	−0,69	0,61	−0,95
	13	0,59	−0,51	0,75	−0,67		13	0,48	−0,61	0,61	−0,84
14	3	0,90	−12,49	1,11	−21,43		14	0,48	−0,55	0,60	−0,74
	4	0,77	−5,38	0,94	−8,30		15	0,48	−0,50	0,61	−0,67
	5	0,69	−3,13	0,84	−4,72		16	0,48	−0,46	0,60	−0,61
	6	0,63	−2,10	0,79	−3,07		17	0,48	−0,44	0,61	−0,57
	7	0,60	−1,50	0,75	−2,16		18	0,48	−0,41	0,61	−0,54
	8	0,58	−1,15	0,73	−1,67	20	3	0,99	−15,33	1,25	−26,67
	9	0,56	−0,93	0,72	−1,30		4	0,83	−6,64	1,02	−10,49
	10	0,56	−0,76	0,72	−1,07		5	0,73	−4,00	0,89	−5,99
	11	0,56	−0,65	0,72	−0,89		6	0,65	−2,73	0,79	−3,95
	12	0,56	−0,57	0,72	−0,76		7	0,59	−2,04	0,72	−2,91
	13	0,56	−0,51	0,72	−0,68		8	0,55	−1,55	0,68	−2,20
	14	0,56	−0,47	0,72	−0,63		9	0,52	−1,26	0,63	−1,72
15	3	0,89	−13,14	1,12	−23,14		10	0,49	−1,03	0,60	−1,42
	4	0,78	−5,55	0,95	−8,79		11	0,47	−0,88	0,59	−1,19
	5	0,70	−3,35	0,85	−4,88		12	0,46	−0,76	0,59	−1,02
	6	0,64	−2,21	0,78	−3,21		13	0,45	−0,67	0,58	−0,89
	7	0,59	−1,56	0,74	−2,29		14	0,45	−0,59	0,57	−0,80
	8	0,57	−1,20	0,71	−1,72		15	0,45	−0,53	0,57	−0,70
	9	0,56	−0,96	0,69	−1,35		16	0,45	−0,49	0,57	−0,65
	10	0,55	−0,82	0,68	−1,10		17	0,45	−0,46	0,57	−0,60
	11	0,54	−0,70	0,68	−0,96		18	0,45	−0,43	0,57	−0,56
	12	0,54	−0,62	0,68	−0,83		19	0,45	−0,41	0,57	−0,53
	13	0,54	−0,55	0,68	−0,73		20	0,45	−0,40	0,57	−0,53
	14	0,54	−0,50	0,68	−0,66						
	15	0,54	−0,46	0,68	−0,59						
16	3	0,92	−13,55	1,13	−22,72						
	4	0,79	−5,89	0,97	−9,38						
	5	0,70	−3,45	0,85	−5,17						

Таблица 46

Коэффициенты $k'(\varepsilon,\alpha)$ и $k''(\varepsilon,\alpha)$ [95]

n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$		n	r	$\alpha = 0{,}90$		$\alpha = 0{,}95$	
		k'	k''	k'	k''			k'	k''	k'	k''
$\varepsilon = 0{,}90$											
3	3	13,16	1,10	20,93	0,75	12	6	5,31	1,46	6,61	1,26
4	3	13,07	1,16	20,93	0,75		7	4,98	1,47	6,09	1,28
	4	8,39	1,16	11,66	0,87		8	4,75	1,47	5,71	1,28
5	3	12,58	1,18	20,38	0,78		9	4,53	1,46	5,40	1,27
	4	8,48	1,23	11,73	0,97		10	4,37	1,47	5,11	1,27
	5	6,73	1,23	8,66	0,97		11	4,23	1,46	4,88	1,27
6	3	11,74	1,18	18,65	0,73		12	4,07	1,47	4,68	1,28
	4	8,18	1,28	11,39	1,00	13	3	8,16	0,72	12,45	−0,45
	5	6,73	1,29	8,89	1,02		4	6,45	1,31	8,82	0,88
	6	5,83	1,27	7,31	1,02		5	5,75	1,45	7,32	1,20
7	3	11,12	1,18	17,54	0,64		6	5,30	1,48	6,49	1,27
	4	7,89	1,31	10,90	1,04		7	4,96	1,49	6,02	1,30
	5	6,68	1,33	8,44	1,08		8	4,73	1,49	5,63	1,30
	6	5,82	1,32	7,23	1,08		9	4,55	1,49	5,32	1,30
	7	5,25	1,32	6,37	1,08		10	4,37	1,49	5,11	1,30
8	3	10,67	1,13	16,36	0,49		11	4,23	1,49	4,90	1,30
	4	7,79	1,33	10,76	1,04		12	4,09	1,49	4,73	1,30
	5	6,50	1,36	8,62	1,11		13	3,96	1,49	4,51	1,30
	6	5,83	1,36	7,18	1,13	14	3	7,69	0,57	11,56	−0,81
	7	5,31	1,36	6,40	1,12		4	6,17	1,29	8,28	0,83
	8	4,90	1,36	5,84	1,12		5	5,54	1,46	6,96	1,18
9	3	10,21	1,12	15,61	0,42		6	5,12	1,51	6,27	1,28
	4	7,39	1,36	10,26	1,06		7	4,82	1,52	5,75	1,32
	5	6,34	1,41	8,13	1,17		8	4,61	1,52	5,47	1,33
	6	5,67	1,41	7,06	1,19		9	4,45	1,52	5,18	1,33
	7	5,28	1,41	6,46	1,19		10	4,30	1,51	4,94	1,33
	8	4,95	1,40	5,94	1,19		11	4,20	1,51	4,79	1,33
	9	4,66	1,40	5,50	1,19		12	4,09	1,51	4,67	1,33
10	3	9,36	0,99	14,88	0,09		13	3,98	1,51	4,51	1,33
	4	7,17	1,34	9,60	0,99		14	3,85	1,51	4,36	1,33
	5	6,13	1,42	8,02	1,17	15	3	7,23	0,43	10,78	−1,05
	6	5,59	1,43	6,99	1,20		4	5,95	1,26	7,94	0,77
	7	5,18	1,43	6,29	1,21		5	5,36	1,44	6,85	1,15
	8	4,91	1,43	5,83	1,21		6	4,97	1,50	6,19	1,29
	9	4,63	1,42	5,51	1,21		7	4,72	1,52	5,77	1,33
	10	4,41	1,42	5,16	1,21		8	4,57	1,52	5,40	1,34
11	3	9,11	0,90	14,47	−0,09		9	4,40	1,52	5,16	1,35
	4	7,04	1,35	9,98	0,97		10	4,26	1,52	4,95	1,35
	5	6,07	1,43	7,83	1,18		11	4,15	1,52	4,76	1,35
	6	5,52	1,45	6,96	1,24		12	4,08	1,52	4,62	1,34
	7	4,96	1,45	6,34	1,25		13	3,98	1,52	4,51	1,35
	8	4,87	1,45	5,82	1,25		14	3,89	1,51	4,39	1,35
	9	4,63	1,44	5,54	1,25		15	3,77	1,52	4,23	1,35
	10	4,44	1,44	5,23	1,25	16	3	7,07	0,25	10,49	−1,38
	11	4,26	1,45	4,94	1,25		4	5,90	1,23	7,94	0,74
12	3	8,40	0,75	12,96	−0,38		5	5,33	1,45	6,73	1,17
	4	6,60	1,34	9,07	0,95		6	4,98	1,52	6,18	1,30
	5	5,79	1,11	7,35	1,20		7	4,74	1,53	5,81	1,35

Продолжение таблицы 46

n	r	$\alpha = 0{,}90$ k'	$\alpha = 0{,}90$ k''	$\alpha = 0{,}95$ k'	$\alpha = 0{,}95$ k''	n	r	$\alpha = 0{,}90$ k'	$\alpha = 0{,}90$ k''	$\alpha = 0{,}95$ k'	$\alpha = 0{,}95$ k''
					$\varepsilon = 0{,}90$						
16	8	4,56	1,54	5,38	1,37	18	16	3,71	1,57	4,14	1,41
	9	4,38	1,54	5,17	1,37		17	3,66	1,58	4,04	1,41
	10	4,24	1,54	4,97	1,37		18	3,59	1,58	3,97	1,42
	11	4,13	1,54	4,79	1,37	20	3	6,04	−0,19	9,01	−2,27
	12	4,05	1,54	4,65	1,37		4	5,24	1,10	7,00	0,46
	13	3,94	1,54	4,49	1,37		5	4,82	1,42	6,16	1,07
	14	3,87	1,54	4,38	1,37		6	4,55	1,54	5,68	1,30
	15	3,79	1,54	4,26	1,37		7	4,41	1,58	5,33	1,38
	16	3,71	1,54	4,16	1,37		8	4,29	1,59	5,06	1,42
18	3	6,43	0,11	9,64	−1,61		9	4,15	1,60	4,85	1,43
	4	5,52	1,19	7,38	0,66		10	4,05	1,60	4,71	1,44
	5	5,04	1,45	6,40	1,10		11	3,97	1,60	4,57	1,44
	6	4,73	1,53	5,79	1,30		12	3,92	1,60	4,46	1,44
	7	4,55	1,56	5,41	1,37		13	3,84	1,60	4,39	1,44
	8	4,39	1,57	5,15	1,40		14	3,77	1,60	4,29	1,44
	9	4,25	1,58	4,99	1,41		15	3,72	1,60	4,20	1,44
	10	4,14	1,58	4,80	1,41		16	3,67	1,60	4,12	1,44
	11	4,05	1,58	4,63	1,41		17	3,61	1,60	4,04	1,44
	12	3,97	1,58	4,51	1,41		18	3,57	1,60	3,97	1,44
	13	3,91	1,57	4,40	1,41		19	3,52	1,60	3,92	1,44
	14	3,84	1,57	4,30	1,41		20	3,47	1,60	3,84	1,44
	15	3,75	1,57	4,21	1,41						
					$\varepsilon = 0{,}95$						
3	3	17,21	1,64	27,32	1,26	9	7	6,91	1,94	8,40	1,67
4	3	17,55	1,73	27,54	1,38		8	6,39	1,94	7,73	1,68
	4	10,88	1,69	15,06	1,36		9	6,00	1,95	7,09	1,69
5	3	17,36	1,79	28,30	1,44	10	3	14,50	1,91	23,00	1,51
	4	11,14	1,76	15,51	1,45		4	10,12	1,98	13,69	1,70
	5	8,68	1,74	11,14	1,44		5	8,39	1,97	11,00	1,72
6	3	16,66	1,83	26,85	1,48		6	7,50	1,97	9,42	1,71
	4	10,95	1,83	15,32	1,52		7	6,83	1,96	8,29	1,71
	5	8,82	1,81	11,58	1,51		8	6,40	1,96	7,61	1,70
	6	7,53	1,80	9,39	1,50		9	6,01	1,96	7,12	1,71
7	3	16,07	1,87	25,31	1,52		10	5,67	1,96	6,65	1,72
	4	10,80	1,88	14,80	1,59	11	3	14,11	1,93	22,60	1,48
	5	8,84	1,86	11,18	1,59		4	10,03	2,01	14,44	1,73
	6	7,61	1,84	9,40	1,57		5	8,34	2,01	10,39	1,75
	7	6,73	1,85	8,19	1,58		6	7,42	2,00	9,39	1,75
8	3	15.76	1,90	24,57	1,53		7	6,83	1,99	8.42	1,75
	4	10,74	1,91	15,22	1,63		8	6,38	1,99	7,65	1,75
	5	8,78	1,90	11,57	1,63		9	6,04	1,98	7,23	1,75
	6	7,67	1,89	9,43	1,62		10	5,75	2,00	6,73	1,76
	7	6,91	1,89	8,38	1,62		11	5,49	2,00	6,35	1,77
	8	6,29	1,89	7,50	1,62	12	3	13,40	1,91	21,39	1,42
9	3	15,33	1,93	23,80	1,55		4	9,56	2,02	13,27	1,74
	4	10,40	1,96	14,41	1,67		5	8,08	2,03	10,40	1,79
	5	8,59	1,95	11,05	1,69		6	7,22	2,02	9,00	1,79
	6	7,51	1,95	9,46	1,68		7	6,66	2,02	8,08	1,79

Окончание таблицы 46

n	r	α = 0,90		α = 0,95		n	r	α = 0,90		α = 0,95	
		k′	k″	k′	k″			k′	k″	k′	k″
ε = 0,95											
12	8	6,27	2,01	7,49	1,79	16	11	5,44	2,09	6,30	1,89
	9	5,95	2,00	7,06	1,78		12	5,30	2,09	6,06	1,88
	10	5,67	2,01	6,63	1,77		13	5,14	2,09	5,87	1,89
	11	5,47	2,01	6,29	1,78		14	5,02	2,10	5,68	1,90
	12	5,26	2,02	6,00	1,80		15	4,91	2,11	5,53	1,90
13	3	13,11	1,92	20,76	1,44		16	4,80	2,12	5,35	1,92
	4	9,47	2,05	13,09	1,78	18	3	11,18	1,83	17,89	1,11
	5	8,04	2,06	10,25	1,82		4	8,46	2,09	11,66	1,81
	6	7,24	2,05	8,89	1,82		5	7,32	2,14	9,43	1,90
	7	6,68	2,05	8,10	1,82		6	6,61	2,14	8,27	1,93
	8	6,29	2,04	7,43	1,81		7	6,23	2,14	7,45	1,94
	9	6,00	2,04	7,83	1,81		8	5,95	2,14	6,97	1,94
	10	5,70	2,04	6,99	1,81		9	5,72	2,14	6,71	1,94
	11	5,50	2,04	6,66	1,81		10	5,51	2,13	6,38	1,93
	12	5,30	2,04	6,36	1,81		11	5,35	2,13	6,15	1,93
	13	5,12	2,05	5,80	1,83		12	5,23	2,14	5,96	1,93
14	3	12,73	1,92	19,14	1,39		13	5,12	2,14	5,75	1,92
	4	9,10	2,06	12,45	1,77		14	5,01	2,13	5,60	1,93
	5	7,82	2,08	9,93	1,84		15	4,88	2,14	5,46	1,94
	6	7,07	2,08	8,71	1,85		16	4,83	2,14	5,36	1,94
	7	6,53	2,07	7,78	1,85		17	4,74	2,15	5,22	1,95
	8	6,16	2,07	7,30	1,85		18	4,65	2,17	5,11	1,97
	9	5,88	2,07	6,92	1,84	20	3	10,78	1,75	16,96	0,94
	10	5,65	2,06	6,49	1,85		4	8,16	2,09	11,03	1,79
	11	5,48	2,07	6,26	1,85		5	7,09	2,15	9,12	1,92
	12	5,31	2,07	6,06	1,85		6	6,49	2,17	8,11	1,94
	13	5,14	2,07	5,80	1,85		7	6,11	2,17	7,42	1,95
	14	4,97	2,08	5,60	1,85		8	5,88	2,17	6,97	1,96
15	3	12,22	1,88	18,38	1,24		9	5,62	2,16	6,56	1,96
	4	8,90	2,06	11,93	1,77		10	5,43	2,16	6,30	1,95
	5	7,64	2,09	9,79	1,84		11	5,29	2,16	6,09	1,95
	6	6,91	2,09	8,55	1,87		12	5,18	2,16	5,92	1,95
	7	6,41	2,08	7,90	1,87		13	5,07	2,16	5,79	1,96
	8	6,10	2,08	7,26	1,86		14	4,96	2,16	5,63	1,96
	9	5,81	2,07	6,87	1,86		15	4,86	2,16	5,49	1,96
	10	5,61	2,07	6,50	1,85		16	4,78	2,16	5,37	1,97
	11	5,43	2,07	6,22	1,85		17	4,70	2,17	5,23	1,98
	12	5,31	2,07	6,00	1,86		18	4,64	2,18	5,14	1,98
	13	5,17	2,07	5,88	1,87		19	4,57	2,18	5,06	1,99
	14	5,02	2,07	5,66	1,87		20	4,49	2,19	4,95	2,00
	15	4,88	2,08	5,46	1,89						
16	3	11,98	1,85	18,76	1,13						
	4	8,88	2,06	12,30	1,77						
	5	7,67	2,11	9,72	1,87						
	6	6,92	2,11	8,63	1,89						
	7	6,47	2,11	7,98	1,89						
	8	6,12	2,10	7,27	1,89						
	9	5,84	2,10	6,92	1,88						
	10	5,60	2,09	6,60	1,88						

Таблица 47

Значения $\eta = t(r,n)/t(r,r)$

i	k							
	1	2	3	4	5	6	7	8
1	1,000							
2	0,500	1,000						
3	0,333	0,556	1,000					
4	0,250	0,389	0,590	1,000				
5	0,200	0,300	0,427	0,616	1,000			
6	0,167	0,244	0,336	0,450	0,635	1,000		
7	0,143	0,206	0,278	0,365	0,479	0,650	1,000	
8	0,125	0,179	0,237	0,365	0,387	0,497	0,663	1,000
9	0,111	0,157	0,207	0,262	0,327	0,406	0,513	0,673
10	0,100	0,141	0,183	0,230	0,283	0,345	0,423	0,526
11	0,091	0,127	0,165	0,205	0,250	0,301	0,361	0,437
12	0,083	0,116	0,150	0,185	0,224	0,267	0,316	0,375
13	0,077	0,107	0,137	0,169	0,202	0,240	0,282	0,330
14	0,071	0,099	0,126	0,155	0,185	0,218	0,254	0,295
15	0,067	0,092	0,117	0,143	0,170	0,200	0,232	0,267

i	k						
	9	10	11	12	13	14	15
1							
2							
3							
4							
5							
6							
7							
8							
9	1,000						
10	0,682	1,000					
11	0,537	0,690	1,000				
12	0,449	0,547	0,696	1,000			
13	0,388	0,460	0,536	0,703	1,000		
14	0,342	0,399	0,470	0,564	0,708	1,000	
15	0,307	0,353	0,409	0,479	0,572	0,713	1,000

Задача 63. *Получена выборка из распределения Вейбулла с параметром* $\beta = 3{,}1$:

$$x_i\colon\ 0{,}74;\ 0{,}78;\ 0{,}94;\ 1{,}26;\ 1{,}39;\ 2{,}17;\ 2{,}19;\ 3{,}18;\ 5{,}16;\ 6{,}12.$$

Необходимо найти оценку максимального правдоподобия параметра α.

Имеем

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\beta}\right)^{\frac{1}{\beta}} = \left[\frac{1}{10}\left(0{,}74^{3,1} + 0{,}78^{3,1} + \ldots + 6{,}12^{3,1}\right)\right]^{\frac{1}{3,1}} = 3{,}53601.$$

Задача 64. *Найти совместные точечные оценки параметров распределения Вейбулла по выборке данных задачи* 63.

Используем метод ускоренного решения системы уравнений максимального правдоподобия. Вычислим оценки среднего и стандартного отклонения

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 2{,}393;$$

$$s = \left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{\frac{1}{2}} = 1{,}88539$$

и найдем оценку коэффициента вариации $\nu = \dfrac{s}{\bar{x}} = 0{,}78777$.

Из соотношения $\beta_0 \approx \nu^{-1{,}075}$ находим приближение

$$\beta_0 = (0{,}787877)^{-1{,}075} = 1{,}292.$$

Далее вычисляем

$$s_1 = \sum_{i=1}^{10} x_i^{1{,}292} = 0{,}74^{1{,}292} + \ldots + 6{,}12^{1{,}292} = 33{,}85538;$$

$$s_2 = \sum_{i=1}^{10} \ln x_i = \ln 0{,}74 + \ldots + \ln 6{,}12 = 6{,}11698;$$

$$s_3 = \sum_{i=1}^{10} x_i^{1{,}292} \cdot \ln x_i = 0{,}74^{1{,}292} \cdot \ln 0{,}74 + \ldots + 6{,}12^{1{,}292} \cdot \ln 6{,}12 = 42{,}26807;$$

$$s_4 = \sum_{i=1}^{10} x_i^{\beta_0} (\ln x_i)^2 = 0{,}74^{1{,}292} \cdot (\ln 0{,}74)^2 + \ldots + 6{,}12^{1{,}292} \cdot (\ln 6{,}12)^2 = 66{,}16118.$$

Вычисляем оценки:

$$\hat{\beta} = \left\{\frac{42{,}26807}{33{,}85538} - \frac{611698}{10} + 1{,}292 \cdot \frac{42{,}26807^2 - 33{,}85538 \cdot 66{,}16118}{33{,}85538^2} \times \right.$$
$$\left. \times \left[\left(\frac{42{,}26807}{33{,}85538} - \frac{6{,}11698}{10}\right) \cdot 1{,}292 - 1\right]\right\}^{-1} = 1{,}37481;$$

$$\hat{\alpha} = \left(\frac{1}{10}\sum_{i=1}^{10} x_i^{1{,}37481}\right)^{\frac{1}{1{,}37481}} = 2{,}58882.$$

Задача 65. *Имеется выборка из распределения Вейбулла:*

$$x_i\text{: } 12,\ 14,\ 21,\ 38,\ 42,\ 44,\ 46,\ 59,\ 61,\ 72.$$

Вычислить оценки параметров α и β методом моментов.

Воспользуемся простейшим методом (оценка по коэффициенту вариации). Вычисляем последовательно

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 40{,}8; \quad s = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{\frac{1}{2}} = 19{,}208; \quad \nu = \frac{s}{\bar{x}} = 0{,}4708.$$

Имеем $\hat{\beta} = 0{,}4708^{-1{,}075} = 2{,}247$ или из табл. 39 $\hat{\beta} \approx 2{,}19$.

Оценку α ищем в форме

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{2{,}247}\right)^{\frac{1}{2{,}247}} = 46{,}0656.$$

Воспользовавшись близостью $\beta \approx 2$, продемонстрируем оценку параметров распределения Рэлея (см. раздел 2.3.1.6). В нашем случае $s = 19{,}208$ и $\bar{x} = 40{,}8$.

Тогда имеем

$$\hat{a}_M = \frac{2s}{\sqrt{4-\pi}} = 41{,}46; \quad \hat{\mu}_M = 40{,}8 - \frac{19{,}208\sqrt{\pi}}{\sqrt{4-\pi}} = 4{,}05.$$

Дисперсии этих оценок приближенно равны ($k_1 = 20{,}745$; $k_2 = 36{,}8886$; $k_3 = 11{,}6404$):

$$D(\hat{a}_M) = \frac{4}{4-\pi}k_1 = 96{,}478 \; (\sqrt{D(\hat{a}_M)} = 9{,}82);$$

$$D(\hat{\mu}_M) = k_2 + \frac{\pi}{4-\pi}k_1 - \frac{2k_3\sqrt{\pi}}{\sqrt{4-\pi}} = 88{,}829 \; (\sqrt{D(\hat{\mu}_M)} = 9{,}42).$$

Полагая (весьма условно — только для демонстрации техники вычислений), что $x_{p_1} = x_{0{,}93\cdot 10} \approx x_9$ и $x_{p_2} = x_{0{,}07\cdot 10} \approx x_1$, имеем для квантильных оценок ($\alpha_1 = -\ln(1-p_1) = 2{,}659$; $\alpha_2 = -\ln(1-p_2) = 0{,}0725$):

$$\hat{a}_k = \frac{x_9 - x_1}{\alpha_1 - \alpha_2} = \frac{61-12}{2{,}659 - 0{,}0725} = 18{,}94;$$

$$\hat{\mu}_k = \frac{\alpha_1 \cdot x_1 - \alpha_2 \cdot x_2}{\alpha_1 \cdot \alpha_2} = \frac{12 \cdot 2{,}659 - 61 \cdot 0{,}0725}{2{,}659 \cdot 0{,}0725} = 10{,}626.$$

Дисперсии их равны:

$$D(\hat{a}_k) = 0{,}7681 \cdot \frac{\hat{a}_k^2}{n} = 27{,}55 \; (\sqrt{D(\hat{a}_k)} = 5{,}25);$$

$$D(\hat{\mu}_k) = 1{,}7506 \cdot \frac{\hat{\mu}_k^2}{n} = 62{,}79 \; (\sqrt{D(\hat{\mu}_k)} = 7{,}92).$$

Не следует приходить в отчаяние от несовпадения оценок $\hat{a}_M$ и $\hat{a}_k$. Это является следствием того, что квантильные оценки применимы при $n > 50$ (у нас, вспомните, речь шла только о демонстрации техники вычислений).

Рассмотрим теперь оценку параметров распределения Вейбулла с помощью коэффициента асимметрии. Отметим сразу, что достаточно точные оценки коэффициента асимметрии требуют объема выборки $n > 100$. Мы преследуем цель продемонстрировать только технику вычислений, поэтому следует помнить, что приведенный пример позволит понять суть метода, но для получения точных оценок с его помощью необходимо располагать значительно бóльшим объемом данных. Имеем

$$\alpha_3 = \frac{1}{10 \cdot s^3} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^3 = \frac{1}{10 \cdot 19{,}208^3} \cdot \left[(12 - 40{,}81)^3 + \ldots + (72 - 4{,}81)^3\right] = -0{,}0707.$$

Примечание. Отличие в оценках от предыдущего метода не должно вводить в заблуждение. Ведь когда мы говорим, что имеем выборку из распределения Вейбулла — это не более, чем гипотеза (мы не знаем даже, сколько в нем значимых параметров). Как узнать, каково же истинное распределение — это самостоятельная задача, рассматриваемая в главе 3 настоящей книги.

Задача 66. *Для данных задачи 65 найти оценки параметров по одной характеристической порядковой статистике.*

Имеем $r' = 0{,}632 \cdot (n + 10) = 0{,}632 \cdot 11 = 6{,}252$; $x_R = \sqrt{x_6 \cdot x_7} = \sqrt{44 \cdot 46} = 44{,}989$.

Вычисляем $G = \dfrac{x_R - \bar{x}}{s} = \dfrac{1}{19{,}208} \cdot (44{,}9889 - 40{,}8) = 0{,}218086$.

Из табл. 41 имеем $\hat{\beta} \approx 1{,}821$ и $b' = 1{,}978$, и окончательно

$$\hat{\alpha} = s \cdot b' = 19{,}208 \cdot 1{,}978 = 37{,}993; \quad \hat{\mu} = x_R - \hat{\alpha} = 44{,}989 - 37{,}993 = 6{,}996.$$

Задача 67. *Имеется гистограмма выборочных данных по наработке приборов*

t_i:	$0 \div 5$	$5 \div 10$	$10 \div 15$	$15 \div 20$	$20 \div 25$	$25 \div 30$	$30 \div 35$	$35 \div 40$
m_i:	21	24	12	10	5	13	9	6

(m_i — *количество данных, попавших в i-й интервал,* $N = 100$). *Найти оценку параметров распределения наработки приборов методом наименьших квадратов (см. раздел* 2.3.1.3).

Вычисления сведем в таблицу (здесь $m_i^* = \frac{1}{N}\left(\sum\limits_{j=1}^{i} m_{ij} - 0{,}5\right)$):

t_i	m_i	m_i^*	P_c	$-\lg P_c$	y	x	x^2	xy
0–5	21	0,205	0,795	0,0996	−1,0016	0,6990	0,4886	−0,70010
5–10	24	0,445	0,555	0,2557	−0,5922	1,0000	1,0000	−0,59220
10–15	12	0,565	0,435	0,3615	−0,4419	1,1761	1,3832	−0,51970
15–20	10	0,665	0,335	0,4749	−0,3233	1,3010	1,6927	−0,42060
20–25	5	0,715	0,285	0,5451	−0,2635	1,3979	1,9541	−0,36847
25–30	13	0,845	0,155	0,8097	−0,0917	1,4771	2,1819	−0,13540
30–35	9	0,935	0,065	1,1871	0,0745	1,5441	2,3841	0,11503
35–40	6	0,995	0,005	2,3010	0,3619	1,6020	2,5666	0,57980
	100				−2,2778	10,1972	13,6512	−2,04154

Находим ($n = 8$ — число интервалов гистограммы):

$$D_0 = n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 = 8 \cdot 13{,}6512 - 10{,}1972^2 = 5{,}2267;$$

$$D_c = \sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} x_i \cdot y_i = (-\,2{,}2778)\ \cdot 13{,}6512 - 10{,}1972 \cdot (-\,2{,}04154) = -\,1{,}2767;$$

$$D_b = n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i = 8 \cdot (-2{,}04154) - 10{,}1972 \cdot (-2{,}2778) = 6{,}89486;$$

$$c = \frac{D_c}{D_0} = -\frac{10{,}2767}{5{,}2267} = -1{,}9661; \quad b = \frac{D_b}{D_0} = \frac{6{,}89486}{5{,}2267} = 1{,}3192.$$

Параметр b является искомой оценкой $\hat{\beta} = b = 1{,}3192$. Параметр a является искомой оценкой $\hat{\alpha}$. Находим ее из условия $c = 0{,}3622 b \lg \alpha$, откуда имеем

$$\lg \hat{\alpha} = -\frac{c + 0{,}3622}{b} = -\frac{-\,1{,}9661 + 0{,}3622}{1{,}3192} = 1{,}215813 \quad \text{и} \quad \hat{\alpha} = 16{,}4366.$$

Задача 68. *В условиях задачи* 65 *найти оценку параметров распределения методом аппроксимации* (*см.* [36]).

Имеем $\alpha_3 = -0{,}0707$ и оценку $\hat{\beta} \approx 48 \cdot (-0{,}0707 + 1{,}23)^{-1{,}4} = 3{,}9027$.
Далее

$$G = \Gamma\left(1 + \frac{1}{\beta}\right) \approx 1 - 0{,}427 \cdot (3{,}9027 - 1) \cdot 3{,}9027^{-1{,}9} = 0{,}6824;$$

$$\delta = s \cdot \left(0{,}5 + 0{,}784 \cdot \hat{\beta} - \frac{0{,}305}{\hat{\beta}}\right) = 19{,}208 \cdot \left(0{,}5 + 0{,}784 \cdot 3{,}9027 - \frac{0{,}305}{3{,}9027}\right) = 66{,}874.$$

Окончательно получаем

$$\hat{\alpha} = \frac{66874}{0{,}6824} = 97{,}998; \quad \hat{\mu} = 40{,}8 - 66{,}874 = -26{,}074.$$

Если принять в качестве исходного двухпараметрическое распределение Вейбулла, то оценки будут равны

$$\hat{\beta} = \frac{10-1}{10} \cdot \left(0{,}465 \cdot \frac{19{,}208}{40{,}8} + 1{,}282 \cdot \frac{40{,}8}{19{,}208} - 0{,}7\right) = 2{,}0178; \quad \hat{\alpha} = \frac{40{,}8}{0{,}6824} = 59{,}79.$$

Напомним, что мы решаем задачу, исходя из того, что закон распределения вероятностей случайной величины нам известен заранее.

Задача 69. *В условиях задачи 67 найти оценки параметров с помощью квантилей.*

Нам потребуются квантили $x_{0,95}$ и $x_{0,2}$.

Имеем $x_{0,95} = x_{[0,05\cdot 100]+1} = x_{51}$; $x_{[0,80\cdot 100]+1} = x_{81}$.

В нашем ряду данных $x_{81} \approx 27$ и $x_{51} \approx 11$, и окончательно получаем $\hat{\alpha} = \exp(0{,}4224 \cdot \ln 27 + 0{,}5776 \cdot \ln 11) = 16{,}073$.

Задача 70. *По данным задачи 65 найти оценки с помощью порядковых статистик (см. раздел 2.3.1.5), если а) цензурированы 5 бо́льших членов выборки; б) выборка не цензурирована.*

Для варианта а) имеем $n = 10$, $r = 5$ и из табл. 42 находим

$$a_1 = -0{,}115524;\quad a_2 = -0{,}090668;\quad a_3 = -0{,}051341;\quad a_4 = 0{,}000925;\quad a_5 = 1{,}256809;$$
$$c_1 = -0{,}185169;\quad c_2 = -0{,}181821;\quad c_3 = -0{,}160697;\quad c_4 = -0{,}125311;\quad c_5 = 0{,}625997.$$

Из табл. 43 находим $k_1 = 0,120330$ и $k_2 = 0{,}177275$. Далее оценки находятся по формулам

$$\hat{u} = \sum_{i=1}^{n} a_i \cdot y_i = -0{,}115524 \cdot \ln 12 \ldots + 1{,}256809 \cdot \ln 42 = 4{,}01772;$$
$$\hat{b} = \sum_{i=1}^{n} c_i \cdot y_i = -0185169 \cdot \ln 12 \ldots + 0{,}652997 \cdot \ln 42 = 0{,}555648.$$

Искомые оценки равны $\hat{\beta} = \frac{1}{\hat{b}} = 1{,}800$; $\hat{\alpha} = \exp(4{,}01772) = 55{,}574$ или (с учетом смещения)

$$\hat{b}^* = \frac{\hat{b}}{1-k_2} = \frac{0{,}555648}{1-0{,}177275} = 0{,}6754;\quad \hat{\beta} = \frac{1}{\hat{b}^*} = 1{,}480;$$
$$\hat{u}^* = \hat{u} + \hat{b}^* \cdot k_1 = 4{,}01772 + 0{,}6754 \cdot 0{,}12033 = 4{,}099;\quad \hat{\alpha}^* = \exp(4{,}099) = 60{,}279.$$

Для случая б) по аналогии с помощью табл. 42 и 43 имеем

$$\hat{u} = 0{,}02331 \cdot \ln 12 + \ldots 0{,}230001 \cdot \ln 72 = 3{,}8619;$$
$$\hat{b} = -0{,}072734 \cdot \ln 12 - \ldots + 0{,}324597 \cdot \ln 72 = 0{,}446372.$$

Из табл. 43 имеем $k_1 = -0{,}020508$ и $k_2 = 0{,}06692$. Тогда

$$\hat{b}* = \frac{0{,}446372}{1-0{,}06692} = 0{,}4784;\quad \hat{u}* = 3{,}8619 + 0{,}4784 \cdot (-0{,}020508) = 3{,}852.$$

Окончательно получаем

$$\hat{\beta} = \frac{1}{\hat{b}^*} = \frac{1}{0{,}4784} = 2{,}090;\quad \hat{\alpha} = \exp(\hat{u}^*) = \exp(3{,}8520) = 47{,}087.$$

Задача 71. *Имеется выборка данных из распределения Вейбулла с параметром*

$$\beta = 2{,}1 - x_i\colon\ 4{,}2;\ 5{,}2;\ 6{,}1;\ 8{,}9;\ 10{,}1;\ 12{,}4;\ 13{,}9;\ 16{,}1;\ 17{,}2;\ 21{,}0.$$

Найти интервальную оценку параметра α при доверительной вероятности $\alpha = 0{,}90$ (Не путать! Разный смысл вкладывается в α — это и параметр распределения и доверительная вероятность. Необходимо следить за указаниями в тексте).

Имеем из табл. 36 для $\alpha = 0{,}90$ и $n = 10$: $k_{\text{н}} = 0{,}637$ и $k_{\text{в}} = 1{,}84$.

Далее $\tilde{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^{\beta} = \frac{1}{10} \cdot \left(4{,}2^{2{,}1} + 5{,}2^{2{,}1} + \ldots + 21^{2{,}1}\right) = 210{,}64941.$

Окончательно

$$\hat{\alpha}_n^{\text{н}} = (0{,}637)^{\frac{1}{2{,}1}} \cdot (210{,}64941)^{\frac{1}{2{,}1}} = 10{,}30338;\quad \hat{\alpha}_n^{\text{в}} = (1{,}84)^{\frac{1}{2{,}1}} \cdot (210{,}64941)^{\frac{1}{2{,}1}} = 17{,}08287.$$

Следовательно, с вероятностью 0,90 значение параметра α находится в диапазоне $10{,}30838 \leqslant \alpha \leqslant 17{,}08287$.

Задача 72. *В условиях задачи* 65 *с учетом точечных оценок параметров, полученных в задаче* 70, *найти интервальные оценки параметров* α *и* β *при доверительной вероятности* $\alpha = 0{,}90$.

Точечные оценки: $\hat{\beta} = 1{,}800$, $\hat{\alpha} = 455{,}574$ при $n = 10$ и $r = 5$ (т. е. когда оценка выполняется по пяти порядковым статистикам) и $\hat{\beta} = 2{,}090$, $\hat{\beta} = 47{,}087$, если оценка ведется по всем $n = r$ наблюдениям. Ищем оценки при $r = 5$ и $n = 10$ ($\alpha = 0{,}90$). Из табл. 44 находим $c' = 0{,}30$ и $c'' = 1{,}53$. Тогда $\hat{\beta}_n^{\text{н}} = 0{,}30 \cdot 1{,}8 = 0{,}54$; $\hat{\beta}_n^{\text{в}} = 1{,}53 \cdot 1{,}8 = 2{,}54$.

Следовательно, $0{,}54 \leqslant \beta \leqslant 2{,}754$.

Для вычисления интервальной оценки α воспользуемся результатами решения задачи 70, из которых следуют оценки $\hat{u}^* = 4{,}099$ и $\hat{b}^* = 0{,}6754$. Из табл. 45 для $r = 5$, $n = 10$ и $\alpha = 0{,}90$ находим $d' = 0{,}72$ и $d'' = -2{,}37$.

Тогда

$$\hat{\alpha}_n^{\text{н}} = \exp(4{,}099 - 0{,}6754 \cdot 0{,}72) = 37{,}066; \quad \hat{\alpha}_n^{\text{в}} = \exp(4{,}099 - 0{,}6754 \cdot (-2{,}37)) = 298{,}777$$
$$\text{и} \quad 37{,}066 \leqslant \alpha \leqslant 298{,}777.$$

По аналогии для случая $r = n$ имеем

$$\hat{\beta} = 2{,}090; \quad \hat{\alpha} = 0{,}71; \quad c' = 0{,}55; \quad c'' = 1{,}38; \quad \hat{u}^* = 3{,}8520; \quad \hat{b}^* = 0{,}4784;$$
$$d' = 0{,}71; \qquad d'' = -0{,}60.$$

Тогда

$$\hat{\beta}_n^{\text{н}} = 0{,}55 \cdot 2{,}090 = 1{,}149; \quad \hat{\beta}_n^{\text{в}} = 1{,}38 \cdot 2{,}090 = 2{,}884;$$
$$\hat{\alpha}_n^{\text{н}} = \exp(3{,}852 - 0{,}4784 \cdot 0{,}71) = 33{,}526; \quad \hat{\alpha}_n^{\text{в}} = \exp(3{,}852 - 0{,}4784 \cdot (-0{,}60)) = 62{,}742.$$

Легко видеть,что увеличение количества наблюдений, участвующих в оценке, повышает ее точность.

Задача 73. *В условиях задачи* 65 *с учетом точечных оценок, полученных в задаче* 70, *найти интервальную оценку* 95%*-й квантили случайной вейбулловской величины* x *при доверительной вероятности* $\alpha = 0{,}90$ *для случая* $r = 5$, $n = 10$.

Для случая $r = 5$, $n = 10$ точечные оценки $\hat{u}^* = 4{,}099$, $\hat{b}^* = 0{,}6754$.

Из табл. 46 находим $k'(0{,}95; 0{,}90) = 8{,}39$ и $k''(0{,}95; 0{,}90) = 1{,}97$.

Тогда имеем

$$\exp(4{,}099 - 0{,}6754 \cdot 8{,}39) \leqslant x_{0,95} \leqslant \exp(4{,}099 - 0{,}6754 \cdot 1{,}97); \quad 0{,}208 \leqslant x_{0,95} \leqslant 15{,}934.$$

Задача 74. *Решить задачу* 73 *для случая* $r = n = 10$.

Для случая $r = n = 10$ имеем

$$\hat{b}^* = 0{,}4784; \quad \hat{u}^* = 3{,}852; \quad k'(0{,}95; 0{,}90) = 5{,}67; \quad k''(0{,}95; 0{,}90) = 1{,}96.$$

Тогда

$$\exp(3{,}852 - 0{,}4784 \cdot 5{,}67) \leqslant x_{0,95} \leqslant \exp(4{,}099 - 0{,}4784 \cdot 1{,}96);$$
$$3{,}125 \leqslant x_{0,95} \leqslant 18{,}436.$$

Задача 75. *Оценить для* $r = 3{,}5$ *и* 7 *при* $n = 10$ *ожидаемую экономию в продолжительности испытаний изделий, если распределение контролируемого параметра подчиняется а) экспоненциальному распределению, б) распределению Вейбулла с параметром* $\beta = 3{,}8$.

Из табл. 47 для случая а) имеем $\eta = 0{,}183$ ($r = 3$); $\eta = 0{,}283$ ($r = 5$) и $\eta = 0{,}423$ ($r = 7$). Таким образом, ожидаемая продолжительность испытаний составит соответственно 18,3%, 28,3% и 42,3% от времени, необходимого при испытаниях в варианте $r = n$. Для случая б) эти величины будут равны $\eta' = 0{,}183^{\frac{1}{3,8}} = 0{,}639 (r = 3)$, $\eta' = 0{,}283^{\frac{1}{3,8}} = 0{,}717$ ($r = 5$); $\eta' = 0{,}423^{\frac{1}{3,8}} = 0{,}797$ ($r = 7$).

2.4. Оценка параметров гамма-распределения

Плотность гамма-распределения имеет вид (см. раздел 1.1.6)

$$f(x) = \frac{1}{\alpha!\beta^\alpha} x^{\alpha-1} \exp\left\{-\frac{x}{\beta}\right\}; \quad x, \beta > 0, \quad \alpha > 1,$$

где α и β — параметры распределения, подлежащие оценке по выборочным данным.

2.4.1. Точечные оценки

2.4.1.1. Оценка β при известном α

$$\hat{\beta} = \frac{\bar{x}}{\alpha}, \quad \text{где} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

2.4.1.2. Совместная оценка параметров

2.4.1.2.1. Оценка максимального правдоподобия

Оценки максимального правдоподобия α и β по выборке объема n являются решением системы уравнений

$$\begin{cases} -n\ln\beta - n\dfrac{\partial[\ln\Gamma(\alpha)]}{\partial\alpha} + \displaystyle\sum_{i=1}^{n}\ln x_i = 0, \\ -n\beta\alpha + \displaystyle\sum_{i=1}^{n} x_i = 0. \end{cases}$$

Система может быть решена методом последовательных приближений. Когда α не слишком мало, можно использовать приближение [16]

$$\frac{\partial}{\partial\alpha}\ln\Gamma(\alpha+1) \approx \ln\left(\alpha + \frac{1}{2}\right) + \frac{1}{24\left(\alpha + \frac{1}{2}\right)^2}.$$

Интересный метод решения системы уравнений максимального правдоподобия предложен в [168–170]. Запишем систему уравнений максимального правдоподобия в виде

$$\begin{cases} \displaystyle\sum_{i=1}^{n}\ln x_i - n\left[\ln\beta + \psi(\alpha)\right] = 0, \\ \displaystyle\sum_{i=1}^{n} x_i - n\beta\alpha = 0, \end{cases}$$

где $\psi(\alpha)$ — логарифм производной гамма-функции.

Предлагается решение системы с помощью параметра

$$A = \ln\frac{\displaystyle\sum_{i=1}^{n} x_i}{n\left(\displaystyle\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}},$$

основанное на аппроксимации зависимости α от A [171].

Из системы уравнений следует $\ln\alpha - \psi(\alpha) = A$. В диапазоне значений $0{,}025 \leqslant A \leqslant 8{,}2$ (что соответствует $0{,}10 \leqslant \alpha \leqslant 20$) применима аппроксимация

$$\hat{\alpha} = \begin{cases} \left(\dfrac{C}{A}\right)^{0.9885} \exp[-0{,}187(C-A)], & A < C; \\ \left(\dfrac{C}{A}\right)^{0.8699}, & A \geqslant C, \end{cases}$$

где $C = 0{,}5772$ — постоянная Эйлера (относительная погрешность аппроксимации не превышает 0,8%).

Параметр β оценивается, как и ранее, по формуле $\hat{\beta} = \dfrac{1}{n\hat{\alpha}} \sum\limits_{i=1}^{n} x_i$.

Оценка для α может быть упрощена:

$$\hat{\alpha} = \begin{cases} 0{,}521143A^{-.09885} \exp(0{,}187A) & \text{при } A < 0{,}5772; \\ 0{,}61998A^{-0.8699} & \text{при } A \geqslant 0{,}5772. \end{cases}$$

Еще одна оценка, основанная на аппроксимации гамма-функции, рассмотрена в [172]. Предлагается разложение гамма-функции по формуле Стирлинга

$$\Gamma(\alpha) = e^{-\alpha}\alpha^{\alpha-\frac{1}{2}}\sqrt{2\pi}\left(1 + \frac{1}{12\alpha} + \frac{1}{288\alpha^2}\right).$$

При использовании первых трех членов разложения приходим к оценке

$$\frac{1}{\hat{\alpha}} = -3 + \sqrt{9 + 12(\ln s_1 - s_2)}, \quad \text{где} \quad s_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s_2 \frac{1}{n}\sum_{i=1}^{n} \ln x_i.$$

Оценка для β в этом случае, как и ранее, определяется по формуле $\hat{\beta} = \dfrac{s_1}{\hat{\alpha}} = \dfrac{\bar{x}}{\hat{\alpha}}$.

2.4.1.2.2. Несмещенная оценка для малых выборок

В [173] предложена оценка

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} x_i \ln x_i - \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \frac{1}{n}\sum_{i=1}^{n} \ln x_i.$$

Ее точность, несмотря на простоту вычислений, вполне удовлетворительна.

2.4.1.2.3. Оценка методом моментов

Метод применим при $n \geqslant 50$. Оценки имеют вид

$$\hat{\alpha} = \left(\frac{\bar{x}}{s}\right)^2; \quad \hat{\beta} = \frac{s^2}{\bar{x}}, \quad \text{где} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

2.4.2. Интервальная оценка параметра β

При известном параметре α интервальная оценка для β имеет вид

$$k_{\text{н}}\bar{x} \leqslant \beta \leqslant k_{\text{в}}\bar{x},$$

где $k_{\text{н}}$, $k_{\text{в}}$ — коэффициенты, которые могут быть взяты из табл. 36 с заменой n на $n\alpha$.

Задача 76. *Имеется ряд выборочных значений случайной величины, имеющей гамма-распределение*

$$x_i: 144, 216, 816, 71, 1147, 2120, 912, 150, 50, 1450, 3500, 189, 21, 914, 1500, 1700, 300, 650.$$

Необходимо найти оценки параметров α и β.

Оценка β при известном α (2.4.1.1)

Пусть $\alpha = 0{,}8$. Тогда оценка равна

$$\hat{\beta} = \frac{\bar{x}}{\alpha} = \frac{1{,}25}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1{,}25}{18} \cdot 15850 = 1100{,}69.$$

Оценка максимального правдоподобия (ускоренный метод) (2.4.1.2.1)

Имеем

$$\sum_{i=1}^{n} x_i = 15850; \quad \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = 431{,}9807; \quad \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = 880{,}555; \quad A = \ln \frac{880{,}555}{431{,}9807} = 0{,}71217.$$

Так как $A > 0{,}5772$, то

$$\hat{\alpha} = 0{,}61998 \cdot 0{,}71217^{-0{,}8699} = 0{,}8329; \quad \hat{\beta} = \frac{\sum_{i=1}^{n} x_i}{n \cdot \hat{\alpha}} = 1057{,}163.$$

Оценка с разложением по формуле Стирлинга (2.4.1.2.1)

Имеем $s_1 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i - 880{,}555$; $s_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \ln x_i = 6{,}06838$.

Далее вычисляем

$$\hat{\alpha} = \frac{1}{-3 + \sqrt{9 + 12 \cdot (\ln 880{,}555 - 6{,}06838)}} = 0{,}84118; \quad \hat{\beta} = \frac{880{,}555}{0{,}84118} = 1046{,}809.$$

Несмещенная оценка (2.4.1.2.2)

Имеем

$$\frac{1}{n} \cdot \sum_{i=1}^{n} x_i \cdot \ln x_i = 6392{,}798; \quad \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = 880{,}555; \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \ln x_i = 6{,}06838.$$

Тогда

$$\hat{\beta} = 6392{,}798 - 880{,}555 \cdot 6{,}06838 = 1049{,}225; \quad \hat{\alpha} = \frac{\bar{x}}{\hat{\beta}} = \frac{880{,}555}{1049{,}225} = 0{,}8392.$$

Оценка методом моментов (2.4.1.2.3)

Имеем

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i - 880{,}555; \quad s^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = 795821{,}914; \quad s = 892{,}0885.$$

Находим

$$\hat{\alpha} = \left(\frac{880{,}555}{892{,}0885}\right)^2 = 0{,}974; \quad \hat{\beta} = \frac{795821{,}914}{880{,}555} = 903{,}773.$$

Отметим, что полученные оценки грубы (существенно отличаются от полученных ранее), так как мал объем выборки (а оценки методом моментов применимы при $n > 50$).

Интервальная оценка β при известном α

Примем, что $\alpha = 0{,}8$ и доверительная вероятность равна 0,95. Имеем $\bar{x} = 880{,}555$, и из табл. 36 для $n \cdot \alpha = 18 \cdot 0{,}8 = 4{,}4$, интерполируя, получаем $k_н \approx 0{,}635$ и $k_в \approx 1{,}805$; окончательно

$$0{,}635 \cdot 880{,}555 \leqslant \beta \leqslant 1{,}805 \cdot 880{,}555, \quad \text{или} \quad 559{,}152 \leqslant \beta \leqslant 1589{,}393.$$

2.5. Оценка параметров биномиального распределения

2.5.1. Точечная оценка

Если имеется реализация из n испытаний, в которых событие наблюдалось m раз, то несмещенной точечной оценкой максимального правдоподобия параметра p является величина $p_n = \dfrac{m}{n}$. Напомним, что плотность вероятностей биномиального распределения имеет вид

$$f(x) = C_n^x p^x (1-p)^{n-x}.$$

Распределение имеет единственный параметр p.

2.5.2. Интервальные оценки

Интервальные оценки параметра p с доверительной вероятностью α являются решениями уравнений Клоппера–Пирсона [174]

$$\sum_{x=m}^{n} C_n^x p_{\text{н}}^x (1-p_{\text{н}})^{n-x} = \frac{1-\alpha}{2}(1-\alpha); \quad \sum_{x=0}^{m} C_n^x p_{\text{в}}^x (1-p_{\text{в}})^{n-x} = \frac{\alpha}{2}(\alpha).$$

В скобках приведены вероятности, соответствующие границам $p_{\text{н}}$ и $p_{\text{в}}$ односторонних доверительных интервалов. Значения $p_{\text{н}}$ и $p_{\text{в}}$, соответствующие различным n и α, приведены в [16, 24, 25, 29].

Известно (см. раздел 1.2.1), что биномиальное распределение может быть аппроксимировано с помощью бета- и F-распределений, нормального распределения и распределения Пуассона. Поэтому значения $p_{\text{н}}$ и $p_{\text{в}}$ для двусторонней интервальной оценки можно выразить через квантили этих распределений.

2.5.2.1. Аппроксимация бета-распределением

$$p_{\text{н}} = B\left(\frac{1-\alpha}{2}; x; n-x+1\right); \quad p_{\text{в}} = B\left(\frac{1+\alpha}{2}; x; n-x+1\right),$$

где $B(\gamma, k, l)$ — γ-квантиль бета-распределения с параметрами k и l.

2.5.2.2. Аппроксимация *F*-распределением

$$p_{\text{н}} = \frac{x}{(n-x+1)F_{\frac{1+\alpha}{2}}[2(n-x+1), 2x]}; \quad p_{\text{в}} = \frac{(x+1)F_{\frac{1+\alpha}{2}}[2(x+1); 2(n-x)]}{n-x+(x+1)F_{\frac{1+\alpha}{2}}[2(x+1); 2(n-x)]},$$

где $F\gamma(k, l)$ — γ-квантиль F-распределения с $f_1 = k$ и $f_2 = l$ степенями свободы.

2.5.2.3. Аппроксимация распределением Пуассона

В данном случае используется связь между распределением Пуассона и распределением хи-квадрат.

$$p_{\text{н}} = \frac{1}{2n}\chi^2_{\frac{1+\alpha}{2}}(2x); \quad p_{\text{в}} = \frac{1}{2n}\chi^2_{\frac{1-\alpha}{2}}[2(x+1)],$$

где χ^2_γ — γ-квантиль распределения хи-квадрат с $f = k$ степенями свободы.

Более точные результаты дает приближение

$$p_{\text{н}} = \frac{\chi^2_{\frac{1+\alpha}{2}}(2x)}{2n - x + 1 + \frac{1}{2}\chi^2_{\frac{1+\alpha}{2}} \pm 2x}; \quad p_{\text{в}} = \frac{\chi^2_{\frac{1-\alpha}{2}}[2(x+1)]}{2n - x + \frac{1}{2}\chi^2_{\frac{1-\alpha}{2}}[2(x+1)]}.$$

При $x = 0$ имеем $p = 0$ и $p_{\text{в}} = 1 - \sqrt[n]{\dfrac{1-\alpha}{2}}$.

Значения $x\left(2n - x + 1 + \frac{1}{2}\chi^2\right)$ и $\frac{1}{n\chi^2}x\left(2n - x + \frac{1}{2}\chi^2\right)$ для разных n, x и α табулированы в [46]. Точностные характеристики этих аппроксимаций исследованы в [175], где показано, что при $1 \leqslant n - x \leqslant 40$, $n \geqslant 40$, при $25 \leqslant n - x \leqslant 40$, $x \geqslant 25$ или при $n - x \geqslant 40$ (x — любое) относительная погрешность не превышает 0,01.

2.5.2.4. Аппроксимация биномиальной суммы распределением хи-квадрат

В [182] предлагается численное решение задачи аппроксимации доверительных границ параметра биномиального распределения, основанное на аппроксимации биномиальной суммы

$$\sum_{i=0}^{x-1} C_n^i q^i (1-q)^{n-i} \approx 1 - P_{2x}\left[2k(q), k(n,x)\right],$$

где $P_{2d}(x)$ — значение функции χ^2-распределения с $2d$ степенями свободы в точке x;

$$k(q) = \ln\frac{1}{1-q}; \quad k(n,x) = \frac{x}{\sum\limits_{i=0}^{x-1}\frac{1}{n-i}}.$$

Отсюда следует

$$p_{\text{н}} = \begin{cases} 1 - \exp\left(-\dfrac{\chi^2_{\frac{1+\alpha}{2}}(2x)}{2k(n,x)}\right), & x \leqslant 0{,}5n; \\ \exp\left\{-\dfrac{\chi^2_{\frac{1-\alpha}{2}}[2(n-x+1)]}{2k(n,n-x+1)}\right\}, & x > 0{,}5n; \end{cases}$$

$$p_{\text{в}} = \begin{cases} 1 - \exp\left\{-\dfrac{\chi^2_{\frac{1-\alpha}{2}}[2(+1)]}{2k(n,x+1)}\right\}, & x \leqslant 0{,}5n; \\ \exp\left\{-\dfrac{\chi^2_{\frac{1+\alpha}{2}}[2(n-x)]}{2k(n,n-x)}\right\}, & x > 0{,}5n. \end{cases}$$

Это наиболее точная аппроксимация из известных. Если n и x достаточно велики, можно использовать приближение

$$k(n,x) \approx \frac{A}{1-A}, \quad \text{где} \quad A = \left(\frac{n-x+1}{n+1}\right)^{\frac{1}{x}}.$$

Погрешность этого приближения равна 0,1% при $n = 10$ и $x = 5$; 0,12% при $n = 20$ и $x = 10$; 0,06% при $n = 40$ и $x = 20$.

2.5.2.5. Аппроксимация нормальным распределением

Для упрощения записи впредь вместо $u_{\frac{1+\alpha}{2}(\alpha)}$ будем использовать обозначение u, т. е. помнить, что u — это $\left(\frac{1+\alpha}{2}\right)$-квантиль (для двустороннего интервала) или α-квантиль (для односторонних интервалов) стандартного нормального распределения (напомним, что α — это доверительная вероятность, с которой определяются интервалы оценки параметра p).

Аппроксимация 1:

$$p_{\text{н}} = \frac{x + \frac{1}{2}u^2 - u\sqrt{\frac{x}{n}(n-x) + \frac{1}{4}u^2}}{n+u^2}; \quad p_{\text{в}} = \frac{x + \frac{1}{2}u^2 + u\sqrt{\frac{x}{n}(n-x) + \frac{1}{4}u^2}}{n+u^2}.$$

Эта аппроксимация рекомендуется [2] при $x \geqslant 4$ и $n - x \geqslant 4$.

Аппроксимация 2 (arcsin):

если биномиальное распределение нормализуется с помощью преобразования arcsin, то

$$p_{\text{н}} = \sin^2\left[\arcsin\sqrt{\frac{x}{n}} - \frac{1}{2\sqrt{n}}u\right]; \quad p_{\text{в}} = \sin^2\left[\arcsin\sqrt{\frac{x}{n}} + \frac{1}{2\sqrt{n}}u\right].$$

Аппроксимация 3 [176]:

$$p_{\text{н}} = \frac{x-0{,}5}{n} - \frac{u}{\sqrt{n}}\sqrt{\frac{x-0{,}5}{n}\left(1 - \frac{x-0{,}5}{n}\right)};$$

$$p_{\text{в}} = \frac{x+0{,}5}{n} + \frac{u}{\sqrt{n}}\sqrt{\frac{x+0{,}5}{n}\left(1 - \frac{x+0{,}5}{n}\right)}.$$

Аппроксимация 4 [176]:

$$p_{\text{н}} = \frac{x - 0{,}5 + \frac{1}{2}u^2 - u\sqrt{x - 0{,}5 - \frac{1}{n}(x-0{,}5)^2 + \frac{1}{4}u^2}}{n+u^2};$$

$$p_{\text{в}} = \frac{x + 0{,}5 + \frac{1}{2}u^2 + u\sqrt{x + 0{,}5 - \frac{1}{n}(x+0{,}5)^2 + \frac{1}{4}u^2}}{n+u^2}.$$

Аппроксимация 5 (при $p \approx 0{,}5$) [175]:

$$p_{\text{н}} = \frac{x-0{,}5}{n} - \frac{u}{\sqrt{n-u^2}}\sqrt{\frac{x-0{,}5}{n}}\left(1 - \frac{x-0{,}5}{n}\right);$$

$$p_{\text{в}} = \frac{x+0{,}5}{n} + \frac{u}{\sqrt{n-u^2}}\sqrt{\frac{x+0{,}5}{n}}\left(1 - \frac{x+0{,}5}{n}\right).$$

Аппроксимация 6 (*Холла*):

$$p_{\text{н}} = \frac{x}{n} - \frac{u}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1 - \frac{x}{n}\right)} + \frac{u^2+0{,}5}{3n}\left(1 - \frac{2x}{n}\right);$$

$$p_{\text{в}} = \frac{x}{n} + \frac{u}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1 - \frac{x}{n}\right)} + \frac{u^2+0{,}5}{3n}\left(1 - \frac{x}{n}\right).$$

Более точная аппроксимация с поправкой 0,5 на непрерывность:

$$p_{\text{н}} = \frac{x-0{,}5}{n} - \frac{u}{\sqrt{n}}\sqrt{\frac{x-0{,}5}{n}\left(1-\frac{x-0{,}5}{n}\right) + \frac{u^2+0{,}5}{3n}\left(1-2\frac{x-0{,}5}{n}\right)};$$

$$p_{\text{в}} = \frac{x+0{,}5}{n} + \frac{u}{\sqrt{n}}\sqrt{\frac{x+0{,}5}{n}\left(1-\frac{x+0{,}5}{n}\right)} + \frac{u^2+0{,}5}{3n}\left(1-2\frac{x+0{,}5}{n}\right).$$

Аппроксимация 7 (*Моленара*):

$$p_{\text{н}} = \frac{(x-1)\left(1+\dfrac{1-u^2}{3n}\right) + \dfrac{2+u^2}{3} - u\sqrt{\dfrac{x(n-x+1)}{n}\left(1+\dfrac{7-u^2}{18n}\right)^2 - (n+1)\dfrac{7-u^2}{18n}}}{n+\dfrac{2+u^2}{3}};$$

$$p_{\text{в}} = \frac{x\left(1+\dfrac{1-u^2}{3n}\right) + \dfrac{2+u^2}{3} + u\sqrt{\dfrac{(x+1)(n-x)}{n}\left(1+\dfrac{7-u^2}{18n}\right) - (n+1)\dfrac{7-u^2}{18n}}}{n+\dfrac{2+u^2}{3}}.$$

Большей точностью обладает аппроксимация

$$p_{\text{н}} = \frac{(x-1)\left(1+\dfrac{1-u^2}{3n}\right) + \dfrac{2+u^2}{3} + \dfrac{1-u^2}{6n} - u\sqrt{\dfrac{x(n-x+1)}{n}\left(1+\dfrac{7-u^2}{18n}\right) - (n+1)\dfrac{7-u^2}{18n}}}{n+\dfrac{2+u^2}{3}};$$

$$p_{\text{в}} = \frac{x\left(1+\dfrac{1-u^2}{3n}\right) + \dfrac{2+u^2}{3} + \dfrac{1-u^2}{6n} + u\sqrt{\dfrac{(x+1)(n-x)}{n}\left(1+\dfrac{7-u^2}{18n}\right) - (n+1)\dfrac{7-u^2}{18n}}}{n+\dfrac{2+u^2}{3}}.$$

Аппроксимация 8 (*Закса*):

$$p_{\text{н}} = \frac{x-1+\dfrac{2+u^2}{3} - u\left[\dfrac{\left(x-\dfrac{7-u^2}{18}\right)\left(n-x+1-\dfrac{7-u^2}{18}\right)}{n+11\cdot\dfrac{7-u^2}{18}-4}\right]^{\frac{1}{2}}}{n+2\cdot\dfrac{2+u^2}{3}-1};$$

$$p_{\text{в}} = \frac{x+\dfrac{2+u^2}{3} + u\left[\dfrac{\left(x+1-\dfrac{7-u^2}{18}\right)\left(n-x-\dfrac{7-u^2}{18}\right)}{n+11\cdot\dfrac{7-u^2}{18}-4}\right]^{\frac{1}{2}}}{n+2\cdot\dfrac{2+u^2}{3}-1}.$$

Аппроксимация 9 (*Полсона–Кэмпа–Пратта*)

Полсон [178], используя преобразование Вилсона–Хилферти [61] для распределения хи-квадрат, показал, что

$$\frac{\left(1-\dfrac{2}{9\nu}\right)F^{\frac{1}{3}}-\left(1-\dfrac{2}{9\mu}\right)}{\sqrt{\dfrac{2}{9\nu}F^{\frac{2}{3}}+\dfrac{2}{9\mu}}}\to N(0,1),$$

где F — квантиль распределения Фишера с $f_1=\mu$ и $f_2=\nu$ степенями свободы (см. раздел 1.1.10).

Кэмп [179] использовал этот факт для аппроксимации биномиального распределения, а Пратт [180] применил для получения доверительных интервалов параметра биномиального распределения

$$p_{\text{в}}=\left[1+\left(\frac{x+1}{n-x}\right)^2\times\right.$$
$$\left.\times\left\{\frac{81(x+1)(n-x)-9n-8-3u\sqrt{9(x+1)(n-x)(9n+5-u^2)}+n+1}{81(x+1)^2-9(x+1)(n-x)(2+u^2)+1}\right\}^3\right]^{-1}.$$

При замене $x+1$ на x и u на $-u$ получаем выражение для $p_{\text{н}}$.

2.5.2.6. Аппроксимация Титенко [181]

В [181] сделана попытка найти более точные, чем нормальные, аппроксимации

$$p_{\text{н}}^*=\rho\left\{1+\frac{1}{4}d\rho^2\left[1+\frac{1}{\left(1-d\rho^2\right)}\right]\right\},$$

$$\text{где}\quad \rho=\frac{f}{1-\dfrac{n-x}{x+1}f};\quad d=\frac{(n-x)\,(n+1)}{(x+1)\,(x+2)};\quad f=\left(\frac{1-\dfrac{\alpha}{2}}{C_n^x}\right)^{\frac{1}{x}},$$

или более точно:

$$p_{\text{н}}=p_{\text{н}}^*=+\frac{dQ\rho^4}{4(1-d\rho^2)^2},$$

где

$$Q=\frac{(n-x)^2(x+1)+(3n-x+4)(x+1)^2}{(x+1)(x+4)(n+1)}.$$

Используя упрощенную запись, получим

$$p_{\text{н}}=\rho\left\{1+\frac{1}{4}d\rho^2\left[1+\frac{\varphi}{\left(1-d\rho^2\right)}\right]\right\},\quad \text{где}\quad \varphi=1+\rho\left(\frac{1+\omega}{2\omega}\right)^{1+\omega}\lg n;\quad \omega=\frac{x}{n}.$$

Аналогично для $p_{\text{в}}$

$$p_{\text{в}}=1-\sigma\left\{1+\frac{1}{4}\beta\sigma^2\left[1+\frac{\varphi'}{\left(1-\beta\sigma^2\right)^2}\right]\right\},$$

где

$$\sigma = \frac{g}{1 - \dfrac{x}{n-x+1}g}; \quad \beta = \frac{x(n+1)}{(n-x+1)(n-x+2)}; \quad g = \left(\frac{\dfrac{1-\alpha}{2}}{C_n^{n-x}}\right)^{\frac{1}{n-x}};$$

$$\varphi' = 1 + \sigma(1+\omega)\big[1{,}8 + (1 - 2{,}2\omega)\lg\omega\big]\lg n; \quad \omega = \frac{x}{n}.$$

Эти результаты следуют непосредственно из решения уравнения Клоппера–Пирсона.

Достаточно подробный обзор и анализ интервальных оценок параметра биномиального распределения приведены в [176]. Различные методы графического решения аппроксимационных задач приведены в [2, 126].

Задача 77. *При $n = 20$ испытаниях имели место 6 событий. Найти 95%-й двусторонний доверительный интервал для вероятности появления события в серии независимых испытаний.*

Точечная оценка вероятности равна $m = \dfrac{6}{20} = 0{,}3$. Имеем $\alpha = 0{,}95$, или $\dfrac{1-\alpha}{2} = 0{,}025$; $\dfrac{1+\alpha}{2} = 0{,}975$. Точные табличные значения искомых вероятностей равны $p_{\text{н}} = 0{,}119$ и $p_{\text{в}} = 0{,}543$ (приводятся для последующего сравнения с аппроксимациями).

Аппроксимация F-распределением (2.5.2.2)

$$p_{\text{н}} = \frac{6}{(20-6+1)\cdot F_{0{,}975}[2\cdot(20-6+1); 2\cdot 6]} = \frac{6}{15\cdot F_{0{,}975}(30{,}12)}.$$

Из таблиц или аппроксимаций (см. раздел 1.1.10) находим $F_{0{,}975}(30{,}12) = 2{,}963$. Тогда

$$p_{\text{н}} = \frac{6}{15\cdot 2{,}963} = 0{,}135;$$

$$p_{\text{в}} = \frac{(6+1)\cdot F_{0{,}975}[2\cdot(6+1); (20-6)]}{20-6+(6+1)\cdot F_{0{,}975}[2\cdot(6+1); (20-6)]} = \frac{7\cdot F_{0{,}975}(14{,}28)}{14+7\cdot F_{0{,}975}(14{,}28)}.$$

Имеем (из таблиц или аппроксимаций) $F_{0{,}975} = 2{,}3$ и $p_{\text{в}} = \dfrac{7\cdot 2{,}3}{14+7\cdot 2{,}3} = 0{,}535$.

Окончательно получаем $0{,}135 \leqslant p \leqslant 0{,}535$.

Аппроксимация распределением Пуассона (2.5.2.3)

Имеем табличные значения $\chi^2_{0{,}975}(2\cdot 6 = 12) = 4{,}4$ и $\chi^2_{0{,}025}(2\cdot 7 = 14) = 26{,}1$. Тогда

$$p_{\text{н}} = \frac{1}{2\cdot 20}\cdot 4{,}4 = 0{,}11; \quad p_{\text{в}} = \frac{26{,}1}{40} = 0{,}652;$$
$$0{,}11 \leqslant p \leqslant 0{,}652.$$

Более точная аппроксимация:

$$p_{\text{н}} = \frac{4{,}4}{2\cdot 20 - 6 + \dfrac{1}{2}\cdot 4{,}4 + 1} = 0{,}118; \quad p_{\text{в}} = \frac{26{,}1}{2\cdot 20 - 6 + \dfrac{1}{2}\cdot 26{,}1} = 0{,}555;$$
$$0{,}118 \leqslant p \leqslant 0{,}555.$$

Аппроксимация биномиальной суммы распределением хи-квадрат (2.5.2.4)

Имеем $A = \left(\dfrac{20-6+1}{21}\right)^{\frac{1}{6}} = 0{,}9454647$; $k(20{,}6) = \dfrac{A}{1-A} = 17{,}336747$; $k(20{,}7) = 16{,}762$. Находим из таблиц $\chi^2_{0975}(12) = 4{,}4$ и $\chi^2_{0{,}025}(14) = 26{,}1$.

Окончательно получаем

$$p_{\text{н}} = 1 - \exp\left\{-\frac{4{,}4}{2 \cdot 17{,}336}\right\} = 0{,}119; \quad p_{\text{в}} = 1 - \exp\left\{-\frac{26{,}1}{2 \cdot 16{,}762}\right\} = 0{,}541;$$

$$0{,}119 \leqslant p \leqslant 0{,}541.$$

Аппроксимация нормальным распределением (2.5.2.5)

Аппроксимация 1:
Имеем $U_{\frac{1+\alpha}{2}} = U_{0,975} = 1{,}96$.

Тогда

$$p_{\text{н}} = \frac{6 + 0{,}5 \cdot 1{,}96^2 - 1{,}96 \cdot \sqrt{\frac{6}{20} \cdot (20-6) + \frac{1}{4} \cdot 1{,}96^2}}{20 + 1{,}96^2}; \quad p_{\text{в}} = \frac{6 + 1{,}9208 + 1{,}96 \cdot \sqrt{4{,}2 + 0{,}9604}}{23{,}8416};$$

$$0{,}145 \leqslant p \leqslant 0{,}519.$$

Аппроксимация 2 (arcsin):

$$p_{\text{н}} = \sin^2\left[\arcsin\sqrt{\frac{6}{20}} - \frac{1}{2 \cdot \sqrt{20}} \cdot 1{,}96\right] = 0{,}124; \quad p_{\text{в}} = \sin^2\left[\arcsin\sqrt{\frac{6}{20}} + \frac{1}{2 \cdot \sqrt{20}} \cdot 1{,}96\right] = 0{,}513;$$

$$0{,}124 \leqslant p \leqslant 0{,}513.$$

Аппроксимация 3:

$$p_{\text{н}} = \frac{6-0{,}5}{20} - \frac{1{,}96}{\sqrt{20}} \cdot \sqrt{\frac{6-0{,}5}{20} \cdot \left(1 - \frac{6-0{,}5}{20}\right)} = 0{,}079;$$

$$p_{\text{в}} = \frac{6+0{,}5}{20} + \frac{1{,}96}{20} \cdot \sqrt{\frac{6+0{,}5}{20} \cdot \left(1 - \frac{6+0{,}5}{20}\right)} = 0{,}530;$$

$$0{,}079 \leqslant p \leqslant 0{,}530.$$

Аппроксимация 4:

$$p_{\text{н}} = \frac{6 - 0{,}5 + \frac{1}{2} \cdot 1{,}96^2 - 1{,}96 \cdot \sqrt{6 - 0{,}5 - \frac{1}{20} \cdot (6-0{,}5)^2 + \frac{1}{4} \cdot 1{,}96^2}}{20 + 1{,}96^2} = 0{,}0573;$$

$$p_{\text{в}} = \frac{6 + 0{,}5 + 1{,}9208 - 1{,}96 \cdot \sqrt{6{,}5 - \frac{1}{20} \cdot 6{,}5^2 + 0{,}9604}}{23{,}8416} = 0{,}543;$$

$$0{,}128 \leqslant p \leqslant 0{,}543.$$

Аппроксимация 5:

$$p_{\text{н}} = \frac{6-0{,}5}{20} - \frac{1{,}96}{\sqrt{20 - 1{,}96^2}} \cdot \sqrt{\frac{6-0{,}5}{20} \cdot \left(1 - \frac{6-0{,}5}{20}\right)} = 0{,}0573;$$

$$p_{\text{в}} = \frac{6{,}5}{20} + \frac{1{,}96}{\sqrt{20 - 1{,}96^2}} \cdot \sqrt{\frac{6{,}5}{20} \cdot \left(1 - \frac{6{,}5}{20}\right)} = 0{,}533;$$

$$0{,}057 \leqslant p \leqslant 0{,}533.$$

Аппроксимация 6 (*Холла*):

$$p_{\text{н}} = \frac{6}{20} - \frac{1{,}96}{\sqrt{20}} \cdot \sqrt{\frac{6}{20} \cdot \left(1 - \frac{6}{20}\right)} + \frac{1{,}96^2 + 0{,}5}{3 \cdot 20} \cdot \left(1 - \frac{2 \cdot 6}{20}\right) = 0{,}128;$$

$$p_{\text{в}} = 0{,}3 + 0{,}43827 \cdot 0{,}45826 + 0{,}07236 \cdot 0{,}4 = 0{,}530;$$

$$0{,}128 \leqslant p \leqslant 0{,}530.$$

Более точная аппроксимация:

$$p_{\text{н}} = \frac{6-0{,}5}{20} - \frac{1{,}96}{\sqrt{20}} \cdot \sqrt{\frac{6-0{,}5}{20} \cdot \left(1 - \frac{6-0{,}5}{20}\right)} + \frac{1{,}96^2+0{,}5}{3\cdot 20} \cdot \left(1 - 2\cdot\frac{6-0{,}5}{20}\right) = 0{,}112$$

$$p_{\text{в}} = \frac{6{,}5}{20} + 0{,}43827 \cdot \sqrt{\frac{6{,}5}{20} \cdot \left(1 - \frac{6{,}5}{20}\right)} + 0{,}07236 \cdot \left(1 - 2\cdot\frac{6{,}5}{20}\right) = 0{,}555;$$

$$0{,}112 \leqslant p \leqslant 0{,}555.$$

Аппроксимация 7 (*Моленара*):

$$p_{\text{н}} = \frac{5\cdot\left(1+\dfrac{1{,}96^2}{3\cdot 20}\right) + \dfrac{2+1{,}96^2}{3} - 1{,}96\cdot\sqrt{\dfrac{6\cdot 15}{20}\cdot\left(1+\dfrac{7-1{,}96^2}{18\cdot 20}\right) - 21\cdot\dfrac{7-1{,}96^2}{18\cdot 20}}}{20+\dfrac{2+1{,}96^2}{3}} = 0{,}1194;$$

$$p_{\text{в}} = \frac{6\cdot 0{,}95264 + 1{,}9472 + 1{,}96\cdot\sqrt{\dfrac{7\cdot 14}{20}\cdot 1{,}00877 - 0{,}18424}}{21{,}9472} = 0{,}544;$$

$$0{,}119 \leqslant p \leqslant 0{,}544.$$

Или более точная аппроксимация:

$$p_{\text{н}} = \\ = \frac{5\cdot\left(1+\dfrac{1-1{,}96^2}{3\cdot 20}\right) + \dfrac{2+1{,}96^2}{3} + \dfrac{1-1{,}96^2}{6\cdot 20} - 1{,}96\cdot\sqrt{\dfrac{6\cdot 15}{20}\cdot\left(1+\dfrac{7-1{,}96^2}{18\cdot 20}\right) - 21\cdot\dfrac{7-1{,}96^2}{18\cdot 20}}}{20+\dfrac{2+1{,}96^2}{3}} = \\ = 0{,}1183;$$

$$p_{\text{в}} = \frac{6\cdot 0{,}95264 + 1{,}9472 - 0{,}02368 + 1{,}96\cdot\sqrt{\dfrac{7\cdot 14}{20}\cdot 1{,}00877 - 0{,}18477 - 0{,}18424}}{21{,}9472} = 0{,}5429;$$

$$0{,}1183 \leqslant p \leqslant 0{,}5429.$$

Аппроксимация 8 (*Закса*):

$$p_{\text{н}} = \frac{6 - 1 + \dfrac{2+1{,}96^2}{3} - 1{,}96\cdot\left[\dfrac{\left(6-\dfrac{7-1{,}96^2}{18}\right)\cdot\left(15-\dfrac{7-1{,}96^2}{18}\right)}{20+11\cdot\dfrac{7-1{,}96^2}{18}-4}\right]^{\frac{1}{2}}}{20+2\cdot\dfrac{2+1{,}96^2}{3}-1} = 0{,}116;$$

$$p_{\text{в}} = \frac{6+1{,}9472+1{,}96\cdot\sqrt{\dfrac{6{,}82453\cdot 13{,}8253}{17{,}93013}}}{22{,}8944} = 0{,}5435;$$

$$0{,}116 \leqslant p \leqslant 0{,}5435.$$

Аппроксимация 9 (*Полсона–Кэмпа–Пратта*):

$$p_{\text{в}} = \left[1 + \left(\frac{7}{14}\right)^2 \cdot \left\{\frac{81\cdot 7\cdot 14 - 9\cdot 20 - 8 - 3\cdot 1{,}96\cdot\sqrt{9\cdot 7\cdot 13\cdot(180+5+1{,}96)^2+21}}{81\cdot 7^2 - 9\cdot 7\cdot(2+1{,}96^2)+1}\right\}^3\right]^{-1} = \\ = 0{,}5428;$$

$$p_{\text{н}} = \left[1 + 0{,}16 \cdot \left\{\frac{7102 + 5{,}88 \cdot 383{,}0918}{2601{,}5536}\right\}^3\right]^{-1} = 0{,}1185; \quad 0{,}1185 \leqslant p \leqslant 0{,}5428.$$

Аппроксимация Титенко (2.5.2.6)

Имеем

$$C_n^x = C_{20}^6 = 38760; \quad f = \left(\frac{\frac{1-0{,}95}{2}}{38760}\right)^{\frac{1}{6}} = 0{,}092707; \quad d = \frac{14 \cdot 21}{7 \cdot 8} = 5{,}25;$$

$$\rho = \frac{0{,}092707}{1 - \frac{4}{7} \cdot 0{,}092707} = 0{,}113809;$$

$$p_{\text{н}}^* = 0{,}113809 \cdot \left\{1 + \frac{1}{4} \cdot 5{,}25 \cdot 0{,}113809^2 \cdot \left[1 + \frac{1}{(-5{,}25 \cdot 0{,}113809^2)^2}\right]\right\} = 0{,}118.$$

Более точные оценки равны:

$$Q = \frac{4^2 \cdot 5 + (60 - 6 + 4) \cdot 7^2}{7 \cdot 10 \cdot 21} = 2{,}6; \quad p_{\text{н}} = 0{,}118 + \frac{5{,}25 \cdot 2{,}6 \cdot 0{,}113809^4}{4 \cdot (1 - 5{,}25 \cdot 0{,}113809^2)^2} = 0{,}1186.$$

В случае упрощенной формы записи имеем

$$\omega = \frac{x}{n} = 0{,}3; \quad \varphi = 1 + 0{,}113809 \cdot \left(\frac{1+0{,}3}{2 \cdot 0{,}3}\right)^{1{,}3} \cdot \lg 20 = 1{,}40457;$$

$$g = \left(\frac{\frac{1-0{,}95}{2}}{38760}\right)^{\frac{1}{14}} = 0{,}37960; \quad \beta = \frac{6 \cdot 21}{15 \cdot 16} = 0{,}525; \quad \sigma = \frac{0{,}37960}{1 - \frac{6}{15} \cdot 0{,}37960} = 0{,}44756;$$

$$\varphi' = 1 + 0{,}44756 \cdot (1 + 0{,}3) \cdot \left[1{,}8 + (1 - 2{,}2 \cdot 0{,}3) \cdot \lg 0{,}3\right] \cdot \lg 20 = 2{,}48535.$$

Окончательно получаем

$$p_{\text{н}} = 0{,}113809 \cdot \left\{1 + \frac{1}{4} \cdot 5{,}25 \cdot 0{,}113809^2 \cdot \left[1 + \frac{1{,}40457}{(1 - 5{,}25 \cdot 0{,}113809^2)^2}\right]\right\} = 0{,}11887;$$

$$p_{\text{в}} = 1 - 0{,}44756 \cdot \left\{1 + \frac{1}{4} \cdot 0{,}525 \cdot 0{,}44756^2 \cdot \left[1 + \frac{2{,}48535}{(1 - 0{,}525 \cdot 0{,}44756^2)^2}\right]\right\} = 0{,}504;$$

$$0{,}1189 \leqslant p \leqslant 0{,}504.$$

2.6. Оценка параметров гипергеометрического распределения

Случайной величиной x, имеющей гипергеометрическое распределение, является число дефектных изделий в выборке объема n из партии изделий объема N, содержащей D дефектных изделий. Практический интерес представляет оценка по выборочному контролю числа дефектных изделий в партии.

Доверительный α-интервал для D определяется границами $D_н$ и $D_в$, являющимися целочисленными решениями неравенств

$$\mathbf{P}(x-1, D_н, n) \geqslant \alpha; \quad \mathbf{P}(x-1, D_н+1, n) < \alpha;$$
$$\mathbf{P}(x, D_в, n) \leqslant 1-\alpha; \quad \mathbf{P}(x, D_в-1, n) > 1-\alpha.$$

где $\mathbf{P}(x, D, n)$ — вероятность появления в выборке объема n ровно x дефектных изделий, если в партии изделий их количество равно D.

Так как $\mathbf{P}(x, D_в, n) = 1 - \mathbf{P}(n-x-1, N-D_в, n)$, то $N - D_в$ является нижним доверительным пределом для $N - D$, построенным по случайной величине $n - x$, и тогда достаточно иметь решение неравенств для $D_н$ при различных α, N, n и x. Таблицы таких решений для различных α, n, $N - n$ и x приведены в [25]. Они составлены для n, удовлетворяющих неравенству $2n \geqslant N$. При $2n < N$ таблицей можно пользоваться со следующей заменой переменных:

$$\tilde{n} = N - n; \quad \tilde{N} - \tilde{n}; \quad \tilde{D}_н - \tilde{x} = n - x.$$

Для заданных $\tilde{n}, \tilde{N} - \tilde{n}$ по таблицам [25] находится значение, соответствующее значению $\tilde{D}_н - \tilde{x} = n - x$, а затем и сама оценка $D_н = x + N - n - \tilde{x}$.

Верхний $\alpha \cdot 100\%$-й доверительный интервал для D по выборке объема n из партии изделий объема N при $x = 0$

$$D \leqslant \left\{1 - (1-\alpha)^{\frac{1}{n}}\right\} \left[N - \frac{n-1}{2}\right],$$

где $[y]$ — целая часть числа y.

Задача 78. *В выборке объема $n = 10$ из партии объема $N = 200$ изделий не обнаружено дефектных изделий ($x = 0$). Найти верхнюю доверительную границу количества дефектных изделий в партии (D) при доверительной вероятности $\alpha = 0{,}95$.*

Имеем

$$D \leqslant \left\{1 - (1-0{,}95)^{\frac{1}{10}}\right\} \cdot \left[200 - \frac{10-1}{2}\right] = 0{,}258865 \cdot 195 = 50.$$

2.7. Оценки при неизвестном законе распределения вероятностей

2.7.1. Оценки для центра распределения

В качестве первичных (достаточно грубых) оценок центра группирования значений случайных величин при неизвестном законе распределения вероятностей могут быть использованы различные предельные неравенства.

2.7.1.1. Неравенства чебышевского типа

2.7.1.1.1. Неравенство Чебышева

Неравенство Чебышева имеет вид

$$\mathbf{P}(|x-\mu| \geqslant k\sigma) < \frac{1}{k^2},$$

где μ и σ — соответственно среднее значение и стандартное отклонение.

Из неравенства Чебышева следует, что

$$x - \frac{\sigma}{\sqrt{1-\alpha}} \leqslant \mu \leqslant x + \frac{\sigma}{\sqrt{1-\alpha}},$$

где α — доверительная вероятность.

Если вместо значения случайной величины x используется выборочное среднее $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$, то имеет место неравенство

$$\bar{x} - \frac{\sigma}{\sqrt{n(1-\alpha)}} \leqslant \mu \leqslant \bar{x} + \frac{\sigma}{\sqrt{n(1-\alpha)}}.$$

Если известно, что распределение симметрично относительно центра μ, то доверительный интервал равен

$$x - \frac{2\sigma}{3\sqrt{1-\alpha}} \leqslant \mu \leqslant x + \frac{2\sigma}{3\sqrt{1-\alpha}} \quad \text{или} \quad \bar{x} - \frac{2\sigma}{3\sqrt{n\,(1-\alpha)}} \leqslant \mu \leqslant \bar{x} + \frac{2\sigma}{3\sqrt{n\,(1-\alpha)}}.$$

Отсюда легко видеть, что только знание того факта, что распределение случайной величины симметрично, уже позволяет построить более узкий доверительный интервал для центра распределения.

2.7.1.1.2. Неравенство Кантелли

$$\mathbf{P}(x-\mu \leqslant \lambda) \leqslant \frac{\sigma^2}{\sigma^2+\lambda^2}, \quad \lambda \geqslant 0.$$

Отсюда следует $\mu \leqslant x - \sigma\sqrt{\dfrac{1-\alpha}{\alpha}}$ или $\mu \geqslant \bar{x} - \sigma\sqrt{\dfrac{1-\alpha}{n\alpha}}$.

2.7.1.1.3. Неравенство Мейделя

Если распределение x имеет единственный максимум в точке μ_o, причем $\tau = \sigma^2 + (\mu-\mu_0)^2$, $s = \dfrac{|\mu-\mu_0|}{\sigma}$, то имеет место неравенство

$$\mathbf{P}\big(|x-\mu_0| > \lambda\tau\big) \leqslant \begin{cases} 1 - \dfrac{\lambda}{\sqrt{3}}, & \lambda \leqslant \dfrac{2}{\sqrt{3}}; \\ \dfrac{4\lambda^2}{9}, & \lambda > \dfrac{2}{\sqrt{3}}, \end{cases}$$

откуда

$$x - \sqrt{3}(1-\alpha)\tau \leqslant \mu \leqslant x + \sqrt{3}\alpha\tau; \quad \lambda \leqslant \frac{2}{\sqrt{3}}$$

$$\text{или} \quad x - \frac{3}{2}\alpha\tau \leqslant \mu \leqslant x + \frac{3}{2}\alpha\tau.$$

2.7.1.2. Оценка Нётера

Нётер [183] показал, что с коэффициентом доверия

$$\alpha = 1 - \left(\frac{1}{2}\right)^{g+h-2} \sum_{i=0}^{g-1} C_{g+h-1}^{i}$$

доверительный интервал для центра симметричного распределения определяется неравенствами

$$\frac{1}{2}\left(x_g + x_{n+1-h}\right) \leqslant \mu \leqslant \frac{1}{2}\left(x_h + x_{n+1-g}\right),$$

где x_i — i-я порядковая статистика.

Значения α для различных g и h табулированы в [183].

При $g + h > 12$ имеет место аппроксимация

$$\alpha \approx 1 - 2\Phi\left(\frac{g-h}{\sqrt{g+h-1}}\right),$$

где $\Phi(x)$ — функция стандартного нормального распределения.

Для распределений нормального типа оптимальная величина $g \approx 0{,}27n$.

Задача 79. *Имеются результаты наблюдений над случайной величиной с неизвестным законом распределения вероятностей (известна только дисперсия $\sigma^2 = 75$):*

x_i: 1,2; 3,4; 6,1; 8,3; 12,1; 13,1; 14,8; 16,7; 21,9; 23,7; 24,5; 28,4.

Найти доверительный интервал для центра распределения при $\alpha = 0{,}95$.

Неравенство Чебышева (2.7.1.1.1)

Имеем

$$\bar{x} = 14{,}516; \quad 14{,}516 - \frac{\sqrt{75}}{\sqrt{12 \cdot (1-0{,}95)}} \leqslant \mu \leqslant 14{,}516 + \frac{\sqrt{75}}{\sqrt{12 \cdot (1-0{,}95)}};$$

$$3{,}336 \leqslant \mu \leqslant 25{,}696.$$

Если бы мы располагали информацией о том, что распределение вероятностей случайной величины x симметрично, то имело бы место

$$14{,}516 - \frac{2\sqrt{75}}{3 \cdot \sqrt{12 \cdot 0{,}05}} \leqslant \mu \leqslant 14{,}516 + \frac{2\sqrt{75}}{3 \cdot \sqrt{12 \cdot 0{,}05}}; \quad 7{,}062 \leqslant \mu \leqslant 21{,}96,$$

т. е. доверительный интервал длины $25{,}696 - 3{,}336 = 22{,}36$ уменьшался бы в 1,5 раза до

$$21{,}96 - 7{,}062 = 14{,}898.$$

Оценка Нётера (2.7.1.2)

Находим $g \approx 0{,}27 \cdot 12 = 3$. Будем использовать для оценки величины $g = 3$ и $h = 5$. Тогда центр распределения находится в интервале

$$\frac{1}{2} \cdot (x_3 + x_8) < \mu < \frac{1}{2} \cdot (x_5 + x_{10}); \quad \frac{1}{2} \cdot (6{,}1 + 14{,}8) < \mu < \frac{1}{2} \cdot (12{,}1 + 23{,}7);$$

$$10{,}45 < \frac{x-\mu}{\sigma} < 17{,}9.$$

Вероятность попадания μ в этот интервал равна

$$\alpha = 1 - \left(\frac{1}{2}\right)^{3+5-2} \cdot \sum_{i=0}^{2} C_7^i = 1 - \left(\frac{1}{2}\right)^6 \cdot (1 + 7 + 21) \approx 0{,}55.$$

2.7.2. Оценка рассеяния распределения

Некоторое представление о степени рассеяния непрерывного распределения дают его выборочные квантили. В общем случае доверительный интервал для p-квантили ограничен элементами упорядоченной по возрастанию выборки с номерами r и s, так как доверительная вероятность равна [5]

$$\alpha = I_p(r, n-r+1) - I_p(s, n-s+1) = \sum_{i=r}^{s-1} C_n^i p^i (1-p)^{n-i} = P(x_r \leqslant x_p \leqslant x_s),$$

где $I_p(a, b)$ — функция бета-распределения с параметрами a и b.

Если $s = n - r + 1$ (случай симметричного интервала), то $\alpha = \sum\limits_{i=r}^{n-r} C_n^i p^i (1-p)^{n-i}$. Значения r и s при $p = 0{,}25$ и $p = 0{,}75$ (т. е. для 25%- и 75%-х квантилей — квартилей) для различных n и α приведены в [46].

Разность между $x_{0,75}$ и $x_{0,25}$, называемая интерквартильной широтой, является характеристикой степени рассеяния распределения относительно его центра.

Задача 80. *В условиях задачи 79 найти доверительный интервал для 25%-й квантили распределения.*

Предположим, что $r = 3$ и $s = n - r + 1 = 10$. Тогда доверительная вероятность того, что в интервале $[x_3 - x_{10}]$ находится 25%-я квантиль ($p = 0{,}25$), равна $\alpha = \sum\limits_{i=3}^{9} C_{12}^i \cdot 0{,}25^i \cdot (1 - 0{,}25)^{12-i} = 0{,}5521664$.

2.8. Некоторые специальные практические задачи

В этом разделе мы рассмотрим ряд задач, встречающихся в практике оценки надежности технических систем. Эти задачи носят специфический характер и не рассматриваются в широко распространенных учебных и рекомендательных пособиях.

2.8.1. Оценка интенсивности отказов с периодом приработки

Известно, что в классическом варианте экспоненциального распределения $f(x) = \lambda \exp(-\lambda x)$ его параметр $\lambda = \text{const}$. На практике обычно интенсивность отказов λ не является постоянной. Начальный этап работы изделия, характеризующийся увеличением интенсивности отказов, называется периодом приработки.

Будем анализировать следующую модель изменения интенсивности отказов во времени:

$$\lambda(t) = \gamma_0 + \gamma_1 e^{-at}.$$

Если время приработки мало (а это обычная практическая ситуация), т. е. когда $\gamma_0, \gamma_1 \ll a$, то для случая, когда n изделий испытываются в течении времени t_n, оценки параметров a, γ_1 и γ_o имеют вид [184]

$$\hat{a} = \frac{1}{t_1} \ln \frac{r_1 - r_2}{r_2 - r_3}; \quad \hat{\gamma}_0 = \frac{1}{t_1} \frac{r_1 r_3 - r_2^2}{n(r_1 - 2r_2 + r_3)}; \quad \hat{\gamma}_1 = \frac{(r_1 - r_2)^3}{n(r_1 - 2r_2 + r_3)^2}\, \hat{a},$$

где $t_1 = \frac{1}{3} t_n$; r_i — количество отказов изделия в промежутки времени $t_1 = t_2 = t_3 = \frac{1}{3} t_n$.

Эффективность таких оценок по отношению к оценкам максимального правдоподобия $\approx 0{,}90$ для $\gamma_0 \approx 0{,}95$, $\gamma_1 \approx 0{,}90 \div 0{,}95$ и $a \approx 0{,}90 \div 0{,}99$.

Задача 81. *20 изделий были испытаны в течение 900 ч. При этом в первые 300 ч получены 4 отказа, во вторые 300 ч — 1 отказ, в последние 300 ч отказов не было. Найти оценку интенсивности отказов с учетом ее изменения во времени.*

Имеем $t_n = 900$; $t_1 = t_2 = t_3 = 300$; $r_1 = 4$; $r_2 = 1$; $r_3 = 0$. Вычислим оценки $\hat{a} = \frac{1}{300} \times \ln \frac{4-1}{1-0} = 3{,}662 \cdot 10^{-3}$; $\hat{\gamma}_0 = \frac{1}{300} \cdot \frac{4 \cdot 0 - 1^2}{4 - 2 \cdot 1 + 0} = 1{,}66 \cdot 10^{-3}$; $\hat{\gamma}_1 = \frac{(4-1)^3}{(4-2+0)^2} = 6{,}75$.

Имеем $\lambda(t) = 1{,}666 \cdot 10^{-3} + 6{,}75 \cdot e^{-3{,}662 \cdot 10^{-3} \cdot t} = 1{,}666 \cdot 10^{-3} + 6{,}75 \cdot (0{,}0134)^t$.

2.8.2. Прогнозирование для экспоненциальных выборок

Для практики представляет интерес следующая задача. Имеются данные по моментам отказов m изделий. Необходимо найти нижний доверительный предел с вероятностью α для среднего времени наработки до отказа для k будущих выборок объема n.

Предположим, что $t_1, t_2, \ldots, t_m$ — наблюдаемые моменты отказа и $\bar{t} = \frac{1}{m} \sum_{i=1}^{m} t_i$. Отберем k выборок объема n. Тогда $\alpha \cdot 100\%$-я доверительная нижняя граница для средних наработок до отказа будущих k выборок есть

$$\bar{t}_{\min} = \bar{t}\omega(k, 2n, 2m, 1-\alpha),$$

где $\omega(k, 2n, 2m, 1-\alpha)$ — коэффициенты оценки.

В [185] для коэффициентов $\omega(\ldots)$ предложена аппроксимация

$$l = \frac{1}{F(2m, 2n; \alpha^{\frac{1}{k}})}.$$

Значения коэффициентов $l' = F(2m, 2n; \alpha^{\frac{1}{k}})$ приведены в табл. 48.

Таблица 48

Значения $l' = F(2m, 2n; \alpha^{\frac{1}{k}})$ для $\alpha = 0{,}95$ [185]

n	$n = 2$					$n = 5$					$n = 10$				
	k					k					k				
	2	3	5	8	12	2	3	5	8	12	2	3	5	8	12
3	0,8	1,1	1,4	1,5	1,7	1,9	2,8	3,6	4,0	4,7	3,3	4,9	6,5	7,9	9,0
5	0,5	0,7	0,9	0,8	0,9	1,1	1,6	2,1	2,5	2,7	2,0	3,0	4,1	4,9	5,6
10	0,2	0,4	0,4	0,4	0,5	0,5	0,8	1,0	1,1	1,3	1,0	1,5	2,0	2,3	2,6
20	0,1	0,2	0,2	0,3	0,2	0,3	0,4	0,5	0,5	0,6	0,5	0,7	0,9	1,0	1,2

$\alpha \cdot 100\%$-ая нижняя граница наработки на отказ k будущих выборок объема n по данным $\bar{t}$ и m равна [186]

$$t_{\min} = \frac{\bar{t}}{kF(2m, 2n; \alpha)},$$

где $F(2m, 2n; \alpha)$ — α-квантиль F-распределения с $f_1 = 2m$ и $f_2 = 2n$ степенями свободы.

Задача 82. *В результате испытаний получены отказы изделий в моменты времени (ч):* $t_1 = 26$, $t_2 = 38$, $t_3 = 110$, $t_4 = 250$, $t_5 = 300$. *Необходимо найти нижнюю 95%-ю границу средних наработок на отказ в пяти будущих выборках изделий объема* $n = 10$ *каждая.*

Имеем $\bar{t} = \frac{1}{5} \cdot \sum_{i=1}^{5} t_i = 144{,}8$. Из табл. 48 для $k = 5$, $n = 10$ и $m = 5$ находим $l' = 4{,}1$. Следовательно, искомая величина равна $\bar{t}_{\min} = \frac{144{,}8}{4{,}1} = 35{,}31$ ч.

Для нижней 95%-й границы наработки на отказ имеем ($F_{0,95}(10{,}20) = 2{,}35$):

$$t_{\min} = \frac{\bar{t}}{5 \cdot F_{0,95}(10{,}20)} = \frac{144{,}8}{5 \cdot 2{,}35} = 12{,}32 \text{ ч.}$$

2.9. Планирование экспериментов для оценки параметров распределений

2.9.1. Нормальное распределение

2.9.1.1. Оценка среднего при известной дисперсии

Объем выборки, необходимый для оценки среднего μ с заданной предельной абсолютной ошибкой ε и доверительной вероятностью α при известной дисперсии σ^2, определяется соотношением

$$n = \left(\frac{u_\alpha \sigma}{\varepsilon}\right)^2,$$

Можно использовать аппроксимацию $u_\alpha \approx 4{,}91[\alpha^{0,14} - (1-\alpha)^{0,14}]$.

Тогда имеем

$$n = 24{,}1081 \left\{\frac{\sigma}{\varepsilon}\left[\alpha^{0,14} - (1-\alpha)^{0,14}\right]\right\}^2.$$

Задача 83. *Напряжение зажигания газоразрядного прибора распределено нормально со стандартным отклонением $\sigma = 50$ В. Найти объем выборки, позволяющий оценить среднее значение напряжения зажигания с предельной абсолютной ошибкой $\varepsilon = 20$ В при доверительной вероятности $\alpha = 0{,}95$.*

Имеем

$$n = 24{,}1081 \cdot \left\{\frac{50}{20} \cdot \left[0{,}95^{0,14} - (1-0{,}95)^{0,14}\right]\right\}^2 = 17.$$

Следовательно, желаемая точность оценки с вероятностью $\geqslant 0{,}95$ достигается при объеме выборки $n \geqslant 17$.

2.9.1.2. Оценка среднего при неизвестной дисперсии

Необходимый объем выборки определяется из соотношения

$$\delta = \frac{\varepsilon}{\bar{x}} = \frac{t_\alpha}{\sqrt{n}}\frac{s}{\bar{x}},$$

где t_α — α-квантиль распределения Стьюдента при $f = n$ степенях свободы; s и $\bar{x}$ — выборочные оценки соответственно стандартного отклонения и среднего значения.

Необходимые значения $\dfrac{t_\alpha(n)}{\sqrt{n}}$ приведены в табл. 49.

Определение объема выборки происходит в следующей последовательности. Сначала по заданным величинам $\delta = \dfrac{\varepsilon}{\bar{x}}$ и α и предполагаемому значению коэффициента вариации $v = \dfrac{s}{\bar{x}}$ по табл. 49 находим значение $\dfrac{t_\alpha(n)}{\sqrt{n}}$ и по нему определяем искомое значение n.

Если для найденного объема выборки n выборочное значение окажется больше предполагавшегося, то эксперимент должен быть продолжен.

Укажем один частный случай, основанный на чрезвычайно простой аппроксимации $t_\alpha(n)$ для $\alpha = 0{,}975$:

$$t_{0,975}(n) = 2\sqrt{\frac{n}{n-2}} \quad \text{и} \quad n = \left(\frac{2s}{\varepsilon}\right)^2 + 2.$$

В этом случае по заданной абсолютной ошибке ε и предполагаемому стандартному отклонению s может быть непосредственно определен объем необходимой

Таблица 49

Значения $t_\alpha(n)/\sqrt{n}$

n	α			n	α			n	α		
	0,90	0,95	0,99		0,90	0,95	0,99		0,90	0,95	0,99
5	0,899	1,150	1,800	25	0,342	0,412	0,558	45	0,250	0,300	0,401
6	0,739	1,000	1,510	26	0,334	0,403	0,545	46	0,248	0,297	0,396
7	0,715	0,890	1,320	27	0,328	0,394	0,532	47	0,245	0,294	0,392
8	0,657	0,816	1,190	28	0,322	0,387	0,521	48	0,242	0,290	0,388
9	0,611	0,754	1,080	29	0,316	0,380	0,513	49	0,240	0,287	0,383
10	0,574	0,706	1,000	30	0,310	0,372	0,502	50	0,237	0,284	0,379
11	0,541	0,663	0,936	31	0,304	0,366	0,492	55	0,226	0,270	0,360
12	0,515	0,630	0,881	32	0,295	0,354	0,475	60	0,216	0,258	0,343
13	0,491	0,598	0,833	33	0,290	0,349	0,468	65	0,207	0,248	0,329
14	0,471	0,572	0,797	34	0,288	0,347	0,463	70	0,199	0,238	0,316
15	0,453	0,550	0,762	35	0,286	0,346	0,461	75	0,192	0,230	0,305
16	0,436	0,530	0,730	36	0,282	0,344	0,459	80	0,186	0,222	0,295
17	0,422	0,512	0,704	37	0,278	0,333	0,447	90	0,175	0,209	0,277
18	0,410	0,495	0,679	38	0,274	0,329	0,441	100	0,166	0,198	0,263
19	0,396	0,479	0,655	39	0,270	0,324	0,434	120	0,151	0,181	0,239
20	0,386	0,466	0,637	40	0,266	0,320	0,428	150	0,135	0,161	0,213
21	0,376	0,454	0,618	41	0,264	0,316	0,422	200	0,117	0,139	0,184
22	0,366	0,442	0,601	42	0,260	0,312	0,417	250	0,104	0,124	0,164
23	0,357	0,431	0,585	43	0,256	0,308	0,411	300	0,095	0,114	0,150
24	0,349	0,421	0,571	44	0,253	0,304	0,406	400	0,082	0,098	0,129

выборки n. Как и ранее, если значение s в эксперименте окажется больше предполагавшегося, то эксперимент должен быть продолжен.

Задача 84. *Определить необходимый объем выборки для оценки среднего значения с предельной относительной ошибкой $\delta = 0{,}4$ при доверительной вероятности $\alpha = 0{,}95$, если предполагаемое значение коэффициента вариации равно $v = 0{,}1$.*

Имеем $\dfrac{t_{0,95}(n)}{\sqrt{n}} = \delta = 0{,}4$. Тогда из табл. 49 для $\alpha = 0{,}95$ непосредственно находим $n = 26$.

Необходимый объем выборки для оценки среднего значения с относительной точностью $\dfrac{\varepsilon}{s} = 0{,}4$ при $\alpha = 0{,}975$ равен $n = \left(2 \cdot \dfrac{1}{0{,}4}\right)^2 + 2 = 27$.

2.9.2. Распределение Вейбулла

Значения необходимых объемов выборок для оценки среднего значения случайной величины, имеющей распределение Вейбулла с известным параметром β, при различных значениях $\delta = \dfrac{\varepsilon}{\bar{x}}$, α и $v = \dfrac{s}{\bar{x}}$ приведены в табл. 50, заимствованной из [44].

Зависимость β от v приведена в табл. 39, из которой по заданному значению β может быть оценено v. Может быть также использована аппроксимация $\nu \approx \beta^{-0{,}9302}$.

Задача 85. *Определить для распределения Вейбулла с параметром $\beta = 1{,}8$ объем выборки, необходимый для оценки среднего значения с относительной погрешностью $\delta = 0{,}15$ при $\alpha = 0{,}95$.*

Имеем $\nu \approx 1{,}8^{-0{,}9302} = 0{,}58$. Из табл. 50 интерполяцией находим $n = 52$. Следовательно, необходимо испытать 52 прибора.

Таблица 50

Значения n для распределения Вейбулла

v	$\alpha = 0,95$					$\alpha = 0,975$				
	δ					δ				
	0,05	0,10	0,15	0,20	0,25	0,05	0,10	0,15	0,20	0,25
0,40	175	45	25	15	10	250	64	33	20	14
0,45	225	60	30	19	13	300	87	42	26	19
0,50	250	75	37	23	17	400	110	53	32	23
0,55	300	90	45	28	20	450	130	64	39	28
0,60	400	113	54	33	23	500	150	76	46	32
0,65	500	138	60	38	27	600	200	87	55	38
0,70	600	150	75	45	32	800	225	110	64	44
0,75	600	175	80	50	36	800	250	120	72	50
0,80	800	200	100	57	38	1000	275	140	80	55
0,85	800	225	113	65	45	1000	300	150	94	64
0,90	800	250	125	75	50	1000	400	175	110	72
0,95	1000	250	138	80	54	1000	400	200	120	76

2.9.3. Биномиальное распределение

Предположим, что задано некоторое значение параметра биномиального распределения — p_0. Тогда наименьший объем выборки, необходимый для того, чтобы подтвердить с вероятностью α, что $p \leqslant p_0$, равен

$$n = \frac{\ln(1-\alpha)}{\ln(1-p_0)}.$$

Если среди n испытанных приборов не будет ни одного отказа, то с вероятностью α можно утверждать, что $p \leqslant p_0$.

Рассмотрим еще одну практическую задачу. Имеем совокупность из N приборов с r дефектными приборами (доля дефектных приборов равна p_0). Необходимо для предполагаемой величины доли дефектных приборов p_0 найти объем выборки n, который с заданной достоверностью α обеспечивает заданную длину доверительного интервала l для оценки p_0. Необходимый объем выборки n в этом случае равен [187]

$$n = \left\{ \frac{l^2}{u^2_{\frac{1+\alpha}{2}} p_0(1-p_0)} + \frac{1}{N} \right\}^{-1}.$$

Для квантили нормального распределения можно использовать аппроксимацию

$$u^2_{\frac{1+\alpha}{2}} = 24{,}1081 \left[\left(\frac{1+\alpha}{2} \right)^{0,14} - \left(\frac{1-\alpha}{2} \right)^{0,14} \right]^2.$$

Приведем еще один результат. Если необходимо найти такой объем выборки n, для которого, при числе дефектных изделий в ней $x = 0$, с вероятностью α можно утверждать, что в партии размера N число дефектных изделий будет не более k, то можно использовать неравенство

$$n \leqslant \left\{ 1 - (1-\alpha)^{\frac{1}{k}} \right\} \left[N - \frac{k-1}{2} \right]$$

где $[\ldots]$ — целая часть числа в скобках.

Задача 86. *Найти объем выборки, позволяющий с достоверностью $\alpha = 0{,}90$ установить, что доля дефектных изделий в партии не превышает заданную величину $P_0 = 0{,}05$.*

Имеем $n = \dfrac{\ln(1-\alpha)}{\ln(1-p_0)} = \dfrac{\ln 0{,}1}{\ln 0{,}95} \approx 45$.

Задача 87. *Необходимо найти объем выборки, при котором для заданной доли дефектных приборов $P_0 = 0{,}1$ в партии из $N = 200$ приборов будет с вероятностью $\alpha = 0{,}95$ получен доверительный интервал для оценки P_0 длиной $l = 0{,}2$.*

Имеем $u_{\frac{1+\alpha}{2}} = u_{0{,}975} = 1{,}96$ и $n = \left\{\dfrac{0{,}04}{1{,}96^2 \cdot 0{,}1 \cdot 0{,}9} + \dfrac{1}{200}\right\}^{-1} \approx 9$.

Задача 88. *Найти объем выборки n, для которого, при отсутствии в выборке дефектных приборов, с достоверностью $\alpha = 0{,}95$ можно утверждать, что в партии $N = 1000$ приборов будет не более 100 дефектных приборов.*

Имеем $n = 1 - \left[(1-0{,}95)^{\frac{1}{100}}\right] \cdot \left[1000 - \dfrac{100}{2}\right] = 28$.

2.9.4. Экспоненциальное распределение

Предположим, что в течение некоторого времени $t_и$ испытывается n приборов и при испытаниях обнаруживается r отказов. Необходимо определить значения n и r, обеспечивающие оценку интенсивности отказов λ с заданной относительной предельной ошибкой δ при доверительной вероятности α.

При испытаниях невосстанавливаемых приборов требуемый объем выборки равен

$$n = \frac{r}{\lambda_0 t_и a},$$

где λ_0 — предполагаемое значение интенсивности отказов; a — коэффициент, зависящий от r и α.

Значения коэффициента $a(r,\alpha)$ приведены в табл. 51. Значения r находятся из соотношения $(1+\delta)^{-1} = b$, где b — коэффициент, зависящий от r и α (его значения приведены в табл. 52). По заданным значениям α и δ сначала определяется b, затем по заданному значению α и вычисленному b из табл. 52 находится r. Далее

Таблица 51

Значения $a(r,\alpha)$

n	α			n	α			n	α		
	0,90	0,95	0,99		0,90	0,95	0,99		0,90	0,95	0,99
1	0,43	0,33	0,22	20	0,77	0,72	0,63	200	0,92	0,89	0,86
2	0,51	0,42	0,30	25	0,79	0,74	0,66	250	0,92	0,90	0,87
3	0,57	0,48	0,36	30	0,80	0,76	0,68	300	0,93	0,91	0,88
4	0,60	0,52	0,40	40	0,83	0,78	0,71	400	0,94	0,92	0,89
5	0,62	0,55	0,43	50	0,84	0,80	0,74	500	0,94	0,93	0,90
6	0,65	0,57	0,46	60	0,86	0,82	0,76	600	0,95	0,94	0,91
8	0,68	0,61	0,50	80	0,87	0,84	0,78	800	0,96	0,94	0,92
10	0,70	0,64	0,53	100	0,88	0,86	0,80	1000	0,96	0,95	0,93
15	0,74	0,68	0,59	150	0,90	0,88	0,84				

Таблица 52

Значения $b(r, \alpha)$

n	α			n	α			n	α		
	0,90	0,95	0,99		0,90	0,95	0,99		0,90	0,95	0,99
1	0,26	0,21	0,15	11	0,66	0,60	0,51	60	0,84	0,81	0,75
2	0,38	0,32	0,24	12	0,67	0,62	0,53	80	0,86	0,83	0,78
3	0,45	0,39	0,30	13	0,68	0,63	0,54	100	0,88	0,85	0,80
4	0,50	0,44	0,35	14	0,69	0,64	0,55	150	0,90	0,87	0,83
5	0,54	0,48	0,38	15	0,70	0,65	0,56	200	0,91	0,89	0,85
6	0,57	0,51	0,40	20	0,74	0,69	0,60	250	0,92	0,90	0,86
7	0,59	0,53	0,44	25	0,76	0,72	0,64	300	0,93	0,91	0,88
8	0,62	0,55	0,46	30	0,78	0,74	0,66	400	0,94	0,92	0,89
9	0,63	0,57	0,48	40	0,81	0,77	0,70	500	0,94	0,93	0,90
10	0,65	0,59	0,50	50	0,83	0,79	0,73	600	0,95	0,94	0,91

для найденного значения r и заданного α по табл. 51 находится значение $a(r, \alpha)$, и по заданному $t_{\text{и}}$ и предполагаемому значению λ_0 вычисляется требуемый объем выборки n. В случае испытаний восстанавливаемых приборов может быть получена оценка необходимого времени испытаний

$$t_{\text{и}} = \frac{r}{a(r, \alpha)} T_0,$$

где T_0 — ожидаемое время наработки на отказ.

Задача 89. *Найти требуемый объем испытаний для оценки интенсивности отказов невосстанавливаемого прибора, если заданы время испытаний* $t_u = 1000$ *ч, предельная относительная ошибка* $\delta = 0{,}2$, *предполагаемое значение интенсивности отказов* $\lambda_0 = 10^{-3}$, *доверительная вероятность* $\alpha = 0{,}95$.

Находим $\dfrac{1}{1+\delta} = \dfrac{1}{1+0{,}2} = 0{,}833$. Из табл. 52 для $b(r, \alpha) = 0{,}833$ и $\alpha = 0{,}95$ находим $r = 80$. Из табл. 51 для $r = 80$ и $\alpha = 0{,}95$ находим $a(r{,}\alpha) = 0{,}84$. Тогда искомый объем выборки $n = \dfrac{80}{10^{-3} \cdot 1000 \cdot 0{,}84} = 95$.

2.9.5. Гамма-распределение

Для оценки среднего значения случайной величины, имеющей гамма-распределение с параметром α, с предельной относительной ошибкой δ при доверительной вероятности γ, объем выборки должен быть не менее

$$n_{\min} = \frac{u_\lambda^2}{(\alpha+1)\delta^2},$$

где u_γ — γ-квантиль стандартного нормального распределения.

Задача 90. *Определить объем выборки, позволяющей найти с предельной относительной ошибкой* $\delta = 0{,}05$ *при доверительной вероятности* $\gamma = 0{,}95$ *среднее значение случайной величины, имеющей гамма-распределение с параметром* $\alpha = 2$.

Имеем искомый объем выборки

$$n = \frac{u_{0,95}^2}{(2+1) \cdot 0{,}05} = \frac{1{,}645^2}{0{,}15} = 18.$$

ГЛАВА 3

МЕТОДЫ АНАЛИЗА ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ СЛУЧАЙНЫХ ВЕЛИЧИН

Общие положения. Для практического применения методов теории вероятностей и математической статистики знание закона распределения вероятностей чрезвычайно важно. По существу, сама изучаемая случайная величина для исследователя представлена только законом распределения вероятностей реализации ее значений.

Зная закон распределения вероятностей наблюдаемой случайной величины, исследователь или инженер в состоянии решать многие практические задачи, связанные с планированием производства, обеспечением качества продукции, оценкой эффективности и стабильности производства.

Попытка применить методы анализа результатов наблюдений, разработанные для конкретных законов распределения вероятностей, в условиях, когда реальное распределение отличается от гипотетического, является самой распространенной на практике ошибкой, приводящей к неверным выводам и, в конечном итоге, к существенным материальным потерям и затратам времени.

Именно поэтому любая обработка результатов наблюдений должна неизменно начинаться с ответа на главный вопрос: каково распределение вероятностей обрабатываемого ряда случайных величин? На практике эта проблема обычно формулируется следующим образом. Выдвигается гипотеза — „наблюдаемое распределение случайных величин описывается некоторым конкретным законом (нормальным, экспоненциальным, Вейбулла, ...)“. Задача первичного исследования — принять или отклонить выдвинутую гипотезу.

Если ни одна из гипотез, связанных с формой закона распределения вероятностей, не принимается, то может быть сформулирована более мягкая гипотеза — например, „наблюдаемое распределение симметрично относительно какой-то точки“. Даже установление только этого факта дает в руки исследователя более эффективные методы анализа наблюдений, чем полное незнание закона распределения вероятностей. И, наконец, если исследователь не получил достаточных оснований для выбора вида распределения, то возникает задача подбора формы распределения непосредственно по экспериментальным данным. При этом распределение вероятностей должно быть подобрано так, чтобы оно удовлетворительно описывало имеющийся экспериментальный материал.

Мы встречаемся здесь с понятием статистической гипотезы. *Статистической гипотезой* называется предположение, выдвигаемое относительно особенностей распределения вероятностей случайной величины, которое проверяется по результатам наблюдений над ней.

Проверка любой статистической гипотезы сводится к следующему. По выборочным значениям случайной величины подсчитывается некоторая величина — *статистический критерий* (*статистика критерия*). При допущении, что распределение вероятностей используемой статистики критерия в условиях справедливости проверяемой гипотезы известно, определяется вероятность появления вычисленного

значения статистики. На основе так называемого принципа значимости устанавливается *уровень значимости* — наибольшее значение вероятности, несовместимое с признанием случайности экспериментально вычисленного значения статистики критерия. Событие называется *значимым* (а не случайным), если теоретическая вероятность его случайного появления меньше, чем принятый уровень значимости. Уровнем значимости определяется критическое значение статистики критерия. Как правило, если значение статистики критерия, вычисленное по экспериментальным данным, больше критического, то гипотеза отклоняется на выбранном уровне значимости. В противном случае она признается не противоречащей результатам наблюдений. Дополнение до единицы уровня значимости называется *уровнем достоверности* (*достоверностью*).

Поскольку статистика критерия для проверки гипотезы вычисляется по выборочным реализациям случайной величины, то и сама она является случайной величиной. Поэтому суждения по гипотезе на основе статистики критерия могут носить только вероятностный характер. При этом различают *ошибки первого рода*, заключающиеся в отклонении верной гипотезы, и *ошибки второго рода*, заключающиеся в принятии ложной гипотезы. Вероятность ошибки первого рода совпадает (по крайней мере не выше) с уровнем значимости и обозначается в литературе через α. Ошибка второго рода обозначается через β. Эффективность статистического критерия проверки гипотезы оценивается его мощностью $1-\beta$, равной вероятности отклонения ложной гипотезы.

Выбор значений α и β определяется условиями эксперимента и требованиями, предъявляемыми к достоверности суждения по проверяемой гипотезе. Обычно на практике используются значения α, β, равные 0,1; 0,05; 0,01.

Проверяемая гипотеза называется нулевой и обозначается символом H_0. Например, запись $H_0\colon F(x) = G(x)$ означает, что проверяется нулевая гипотеза о совпадении функций распределения $F(x)$ и $G(x)$. Подробно теория статистических гипотез изложена в [1, 132, 188–192].

В классификации статистических критериев проверки гипотез о законе распределения вероятностей принята определенная терминология. Такие критерии подразделяются на два класса — общие критерии согласия и специальные критерии согласия. *Общие критерии согласия* применимы к самой общей формулировке гипотезы, как гипотезы о согласии наблюдаемых результатов с любым априорно предполагаемым распределением вероятностей. *Специальные критерии согласия* предполагают специальные нулевые гипотезы, формулирующие согласие с определенной формой распределения вероятностей — нормальной, экспоненциальной, Вейбулла и т. д. Такие критерии носят соответствующие названия — критерии нормальности, критерии экспоненциальности и т. п.

Естественно, что при формулировании специфических требований общие критерии согласия могут быть трансформированы в специальные критерии. Следует отметить, что многообразие возможных альтернатив, противостоящих нулевой гипотезе, порождает и чрезвычайное многообразие статистических критериев, имеющих различную мощность по отношению к различным альтернативам. Поэтому далее приведена весьма большая гамма известных статистических критериев, впервые собранная в одной книге.

3.1. Общие критерии согласия

Нулевая гипотеза при применении общих критериев согласия записывается в форме

$$H_0\colon F_n(x) = F(x),$$

где $F_n(x)$ — эмпирическая функция распределения вероятностей; $F(x)$ — гипотетическая функция распределения вероятностей.

Все известные общие критерии согласия можно разбить на три основные группы:

— критерии, основанные на изучении разницы между теоретической плотностью распределения и эмпирической гистограммой;

— критерии, основанные на расстоянии между теоретической и эмпирической функциями распределения вероятностей;

— корреляционно-регрессионные критерии, основанные на изучении корреляционных и регрессионных связей между эмпирическими и теоретическими порядковыми статистиками.

Кроме критериев, входящих в перечисленные группы, известен ряд критериев, использующих специфичные характеристические свойства различных распределений, и они будут представлены в книге в разделах, посвященных специальным критериям согласия, ориентированным на фиксированные нулевые гипотезы.

3.1.1. Критерии, основанные на сравнении теоретической плотности распределения и эмпирической гистограммы

3.1.1.1. Критерий согласия χ^2

Критерий основан на сравнении эмпирической гистограммы распределения случайной величины с ее теоретической плотностью. Диапазон изменения экспериментальных данных разбивается на k интервалов, и подсчитывается статистика

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i},$$

где n_i — количество значений случайной величины, попавших в i-й интервал; $n = \sum\limits_{i=1}^{k} n_i$ — объем выборки; $F(x)$ — гипотетический теоретический закон распределения вероятностей случайной величины; $p_i = F(x_{i+1}) - F(x_i)$ — теоретическая вероятность попадания случайной величины в i-й интервал.

Дисперсия статистики критерия χ^2 равна [193]

$$\mathbf{D}(\chi^2) = 2(k-1) + \frac{1}{n}\left(\sum_{i=1}^{k} \frac{1}{p_i} - k^2 - 2k + 2\right).$$

Если $\sum\limits_{i=1}^{k} \dfrac{1}{p_i} \ll n$ и $k \ll n$, то $\mathbf{D}(\chi^2) = 2(k-1)$, т. е. совпадает с дисперсией случайной величины, имеющей χ^2-распределение. На этой основе принято считать, что статистика χ^2 имеет распределение, близкое к распределению хи-квадрат (см. раздел 1.1.8).

На мощность статистического критерия χ^2 сильное влияние оказывает число интервалов разбиения гистограммы (k) и порядок ее разбиения (т. е. выбор длин интервалов внутри диапазона изменения значений случайной величины). На практике принято считать, что статистику χ^2 можно использовать, когда $np_i \geqslant 5$.

В [194, 195] показано, что такое приближение допустимо и тогда, когда не более, чем в 20% интервалов имеет место $1 \leqslant np_i \leqslant 5$ (для гладких унимодальных альтернатив). В [196, 197] рекомендуется при $n \geqslant 200$ выбирать k из условия $k = 4\left\{0{,}75(n-1)^2\right\}^{\frac{1}{5}} \approx 3{,}78(n-1)^{\frac{2}{5}}$. Можно рекомендовать еще одно простое правило — нужно выбирать как можно большее k, но не превышающее $n/5$. Исчерпывающие рекомендации по методологии выбора числа k приведены в проекте методических рекомендаций „Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть 1. Критерии типа χ^2. Госстандарт России, 2000 г.“, разработанном Лемешко Б. Ю., Денисовым В. И. и Постоваловым С. Н. из Новосибиского государственного технического университета. Границы интервалов рекомендуется выбирать случайно, исходя из условия $p_i = \dfrac{1}{k} = \text{const}$. Однако в [198] показано, что мощность критерия χ^2 с увеличением числа классов разбиения снижается, и оптимальная мощность соответствует $k \approx 10$.

Укажем также на следующие правила выбора k:

$$k = 1 + 3{,}32\lg n \;\; [314]; \quad k = b\left\{\frac{\sqrt{2}(n-1)}{u_\beta + u_{1-\beta}}\right\}^{\frac{2}{5}} \;\; [11],$$

где α, β — ошибки первого и второго рода; b — коэффициент из диапазона $2 \div 4$ (для простой гипотезы и $\beta \approx 0{,}05$ рекомендуется $b = 4$ [194]); u_γ — γ-квантиль стандартного нормального распределения.

Так или иначе, статистика χ^2 имеет распределение хи-квадрат с $f = n - 1$ степенями свободы в том случае, когда проверяется простая нулевая гипотеза H_0, т. е., когда гипотетическое распределение, на соответствие которому проверяется эмпирический ряд данных, известно с точностью до значения своих параметров. Если гипотеза сложная и параметры гипотетического распределения оцениваются по самой выборке, то число степеней свободы уменьшается на число оцениваемых параметров m и равно $f = k - 1 - m$.

Правило проверки гипотезы просто: если

$$\sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} > \chi^2_\alpha(f),$$

то на уровне значимости α, т. е. с достоверностью $(1-\alpha)$ гипотеза H_0 отклоняется. Вопрос о выборе k с учетом частных альтернатив рассмотрен в [196, 199–201]. Например, в [196] показано, что против альтернативных распределений с „тяжелыми хвостами“ необходимо выбирать k сравнительно большим. В [202] рассмотрена частная задача построения χ^2-критерия со случайными интервалами разбиения для проверки нормальности распределения, когда параметры распределения оцениваются не по группированной выборке ($\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$; $s = \left\{\dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{1}{2}}$). Подробно такой подход рассмотрен в разделе 3.2, посвященном критериям нормальности распределения.

Аппроксимация статистики χ^2 для малых значений p_i и ее поведение изучались в [203, 204] в сравнении с двумя другими критериями максимального правдоподобия

$$Y^2 = 2\sum_{i=1}^{k} n_i \ln\frac{n_i}{np_i}; \quad T^2 = \sum_{i=1}^{k}\left(\sqrt{n_i} + \sqrt{n_{i+1}} - \sqrt{4np_i + 1}\right).$$

(в этом случае разница будет ощущаться на „хвостах“ распределений).

Показано, что критерий χ^2 предпочтительнее указанных критериев. Для случая малых p_i в [205] предложено правило: если $k \geqslant 3$ и число попаданий (ожидаемых) $r < 5$, то можно использовать соотношение $(np_i)_{\min} = 5\frac{r}{n}$.

Если $\frac{1}{n}\sum_{i=1}^{k}\frac{1}{p_i}$ велико (p_i мало), то дисперсия $D(X^2) > D(\chi^2)$, т. е. в этом случае разница будет ощущаться на „хвостах“ распределения. Поэтому в [203] предложена аппроксимация χ^2-критерия с помощью двухпараметрического логарифмически нормального распределения (см. раздел 1.1.3), как это предложено в [206]. Аппроксимация имеет вид $\mathbf{P}(\chi^2 > z) \approx \mathbf{P}(Z > z)$, где z — логнормальная случайная величина с параметрами распределения вероятностей μ и σ:

$$\mu = 2\ln(k-1) - \frac{1}{2}\left\{k^2 - 1 + \frac{1}{n}\left(\sum_{i=1}^{k}\frac{1}{p_i} - k^2 - 2k + 2\right)\right\};$$

$$\sigma^2 = \ln\left\{k^2 - 1 + \frac{1}{n}\left(\sum_{i=1}^{k}\frac{1}{p_i} - k^2 - 2k + 2\right)\right\} - 2\ln(k-1).$$

Если u_α — α-квантиль стандартного нормального распределения, то $\chi^2_\alpha = \exp(\mu + \sigma u_\alpha)$. Мощность χ^2-критерия снижается в связи с тем, что он не учитывает знака разностей $(n_i - np_i)$. Эту информацию использует критерий серий. Его статистикой является число подряд следующих групп „плюсовых“ и „минусовых“ значений $(n_i - np_i)$. Например, для последовательности знаков разностей $(n_i - np_i)$ „$\ldots + + - + - + + - - - + - \ldots$“ число плюсовых серий равно $M = 4$, а число минусовых серий $N = 4$, общее число серий равно $R = M + N = 8$.

Известно, что

$$\mathbf{P}(R = 2x) = \frac{2C_{M-1}^{x-1}C_{N-1}^{x-1}}{C_{M+N}^{M}}; \quad \mathbf{P}(R = 2x-1) = \frac{C_{M-1}^{x-2}C_{N-1}^{x-1} + C_{M-1}^{x-1}C_{N-1}^{x-2}}{C_{M+N}^{N}};$$

$$\mathbf{M}(R) = 1 + \frac{2MN}{M+N}; \quad \mathbf{D}(R) = \frac{2MN(2MN - M - N)}{(M+N)(M+N-1)}.$$

Случайная величина $\frac{R - \mathbf{M}(R)}{\sqrt{\mathbf{D}(R)}}$ распределена приблизительно нормально. Хотя этот критерий имеет мощность меньше, чем χ^2-критерий, но он от него не зависит и поэтому их можно использовать совместно.

Следует помнить, что при сложной гипотезе (т. е., когда параметры гипотетического распределения оцениваются по имеющейся единственной выборке) критерий χ^2 еще асимптотически независим от критерия серий, но в этом случае распределение χ^2 известно лишь приближенно, а распределение критерия серий совсем неизвестно. Таким образом, критерий, комбинирующий критерии χ^2 и серий, применим только для полностью определенного (с точностью до параметров) гипотетического распределения.

Если α_1 — уровень значимости для χ^2-критерия, а α_2 — уровень значимости для критерия серий, то уровень значимости комбинированного критерия будет равен $\alpha = \alpha_1\alpha_2(1 - \ln\alpha_1\alpha_2)$.

Следует иметь в виду, что величина $\alpha = -\ln\alpha_1\alpha_2$ имеет распределение χ^2 с $f = 4$ степенями свободы, что и является, по существу, основанием для построения комбинированного критерия.

В заключение приведем простые (хотя и не менее чувствительные) правила проверки нулевой гипотезы (приближенные критерии):

— если $\chi^2 > f$, то H_0 отклоняется;

— если $R = \dfrac{|\chi^2 - f|}{\sqrt{2f}} \geqslant 3$, то H_0 отклоняется (критерий Романовского [207]).

Задача 91. *Имеем ряд выборочных значений случайной величины* ($n = 100$):

43	76	84	91	95	101	105	114	122	129
54	77	84	91	96	101	106	114	122	132
56	77	85	91	96	101	107	115	122	134
57	78	85	91	96	103	107	116	123	136
61	78	86	92	97	103	107	116	124	136
64	79	87	92	97	104	108	116	124	138
67	79	87	93	98	104	111	117	125	143
73	82	87	93	98	104	112	118	125	143
74	82	88	93	99	104	113	118	125	145
76	83	89	95	101	105	114	119	126	150

Необходимо проверить критерием χ^2 гипотезу о том, что распределение случайной величины не противоречит нормальному закону с параметрами $\mu = 101$ и $\sigma = 16$ на уровне значимости $\alpha = 0{,}1$.

Сначала примем решение, на какое количество классов следует разбить гистограмму эмпирического распределения. Различные рекомендации дают следующие результаты (примем $\alpha = 0{,}1$ и $\beta = 0{,}2$):

$$k = 4 \cdot \left[0{,}75 \cdot (n-1)^2\right]^{\frac{1}{5}} = 4 \cdot (0{,}75 \cdot 0{,}99^2)^{\frac{1}{5}} = 24; \quad k = 1 + 3{,}32 \ln n = 1 + 3{,}32 \ln 100 = 8.$$

Учитывая, что первая рекомендация эффективна при $n \geqslant 200$, и исходя из ограничения $k \leqslant \dfrac{n}{5} = 20$, примем $k = 8$. Продемонстрируем теперь технику вычисления теоретических вероятностей p_i. Пусть x_i и x_{i+1} — границы i-го класса разбиения. Тогда теоретическая вероятность попадания случайной величины в этот интервал равна $F\left(\dfrac{x_{i=1} - \mu}{\sigma}\right) - F\left(\dfrac{x_i - \mu}{\sigma}\right) = p_i$, где $F(x)$ — функция стандартного нормального распределения.

Для нахождения $F(x)$ можно использовать либо таблицы, либо аппроксимации. Приведем одну аппроксимацию [36]

$$F(x) = 1 - 0{,}852 \cdot \exp\left\{-\left(\frac{x + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\}, \quad x > 0.$$

При $x < 0$ используется соотношение $F(-x) = 1 - F(x)$.

Вероятность попадания случайной величины в интервал $x_i < x \leqslant x_{i+1}$, где $x_i = 90$ и $x_{i+1} = 100$, равна

$$p_i = F\left(\frac{100-101}{16}\right) - F\left(\frac{90-101}{16}\right) = F\left(-\frac{1}{16}\right) - F\left(-\frac{11}{16}\right) =$$
$$= 1 - F\left(\frac{1}{16}\right) - 1 + F\left(\frac{11}{16}\right) = F\left(\frac{11}{16}\right) - F\left(\frac{1}{16}\right) = 0{,}229367.$$

Выберем границы классов разбиения из условия равномерного разбиения диапазона изменения случайной величины на 8 классов, с условием попадания в крайние классы не менее 5 наблюдений. Результаты сведем в таблицу:

i	x_i	n_i	$F(x_{i+1})$	$F(x_i)$	p_i	np_i	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
1	< 70	7	0,0263	0,0000	0,0263	2,6300	19,0969	7,2610
2	70–80	10	0,0945	0,0263	0,0682	6,8200	10,1124	1,4830
3	80–90	13	0,2458	0,0945	0,1513	15,1300	4,5369	0,2998
4	90–100	18	0,4751	0,2458	0,2293	22,9300	24,3049	1,0600
5	100–110	17	0,7131	0,0475	0,2380	23,8000	46,2400	1,9428
6	110–120	14	0,8827	0,7131	0,1696	16,9600	8,7616	0,5166
7	120–130	12	0,9650	0,8827	0,0824	8,2400	14,1317	1,7157
8	> 130	9	1,0000	0,9650	0,0350	3,5000	30,2500	8,6428
		100			1,0	10,1972	13,6512	$\chi^2 = 22{,}9217$

Итак, мы получили значение статистики критерия $\chi^2 = 22{,}9217$. Теперь необходимо найти критическое значение статистики, равное $\chi^2_{1-\alpha}$ $(f = k - 1)$.

В нашем случае число степеней свободы равно $f = k - 1 = 8 - 1 = 7$. Используем для вычисления критического значения аппроксимацию Вилсона–Хилферти

$$\chi^2_\alpha(f) = f \cdot \left(1 - \frac{2}{9f} + u_\alpha \sqrt{\frac{2}{9f}}\right)^3,$$

где u_α — α-квантиль стандартного нормального распределения. В нашем случае для $1 - \alpha = 1 - 0{,}1 = 0{,}9$ имеем

$$u_{0,9} = 1{,}28 \quad \text{и} \quad \chi^2_{0,90}(7) = 7 \cdot \left(1 - \frac{2}{9 \cdot 7} + 1{,}28 \cdot \sqrt{\frac{2}{9 \cdot 7}}\right)^3 = 11{,}98.$$

Так как $\chi^2 = 22{,}92 > 11{,}98$, нулевая гипотеза отклоняется, т. е. утверждение о том, что исследуемая выборка взята из нормального распределения с параметрами $\mu = 101$ и $\sigma = 16$, не подтверждается.

Проверим возможность использования логнормальной аппроксимации для χ^2. Находим

$$\sum_{i=1}^{k} \frac{1}{p_i} = 114{,}4613; \quad \mu = 2\ln 7 - \frac{1}{2}\ln\left\{7^2 - 1 + \frac{114{,}4613 - 7^2 - 2 \cdot 7 + 2}{100}\right\} = 1{,}95068.$$

Далее

$$\sigma^2 = \ln\left[49 - 1 + \frac{114{,}4613 - 49 - 14 + 2}{100}\right] - 2\ln 7 < 0.$$

Видим, что аппроксимация в нашем случае неприменима, так как $\sum_{i=1}^{k} \frac{1}{p_i} \approx n$.

Теперь применим комбинированный критерий знаков разности $(n_i - np_i)$ с учетом критерия серий. Последовательность знаков разности $(n_i - np_i)$ в нашем случае имеет вид $+ + - - - - ++$.

Таким образом, всего имеется $R = 3$ серии знаков, в том числе положительных $M = 2$ и отрицательных $N = 1$. Тогда имеем

$$\mathbf{M}(R) = 1 + \frac{2 \cdot 2 \cdot 1}{3} = 2{,}33; \quad \mathbf{D}(R) = \frac{2 \cdot 2 \cdot 1 \cdot (2 \cdot 2 \cdot 1 - 2 - 1)}{(2+1)^2 \cdot (2 + 1 - 1)} = 0{,}222;$$

$$\frac{R - \mathbf{M}(R)}{\sqrt{\mathbf{D}(R)}} = \frac{302{,}33}{\sqrt{0{,}222}} = 1{,}415.$$

Находим $F(1{,}415) = 1 - 0{,}852 \cdot \exp\left\{-\left(\frac{1{,}415 + 1{,}5774}{2{,}0637}\right)^{2,34}\right\} = 0{,}9216.$

Следовательно, уровень значимости $\alpha_1 = 1 - 0{,}9216 = 0{,}0784$.

Для критерия $\chi^2(7)$ уровень значимости равен $\alpha_2 \approx 0{,}0005$ и $\alpha = -2\ln 0{,}0784 \times \times 0{,}0005 = 20{,}29$, что также отклоняет нулевую гипотезу, так как $\chi^2_{0,95}(4) = 9{,}49 < 20{,}29$.

Теперь приведем другой вариант решения задачи, чтобы продемонстрировать технику применения критерия χ^2. Предположим, что параметры распределения неизвестны и определяются по выборке. Примем то же самое число классов разбиения — 8, однако границы классов разбиения будем определять из условия $p_\imath = \dfrac{1}{k} = 0{,}125 = \text{const}$. Тогда, например, для первого класса должно быть $F\left(\dfrac{x_\imath - \bar{x}}{s}\right) = 0{,}125$; $\dfrac{x_i - \bar{x}}{s} = u_{0,125}$; $x_1 = \bar{x} + s \cdot u_{\frac{1}{k}}$. Для второго класса $x_2 = \bar{x} + s \cdot u_{0,25}$, и в общем виде $x_\imath = \bar{x} + s \cdot u_{\frac{1}{k}}$.

Оценки параметров распределения по выборке равны

$$\bar{x} = \frac{1}{n} \cdot \sum_{\imath=1}^{n} x_i = 100{,}77; \quad s = \left\{\frac{1}{n-1} \cdot \sum_{\imath=1}^{n} (x_\imath - \bar{x})\right\}^{\frac{1}{2}} = 21{,}691.$$

Для u_α будем использовать аппроксимацию

$$u_\alpha \approx 4{,}91 \cdot \left\{\alpha^{0,14} - (1-\alpha)^{0,14}\right\}.$$

Результаты расчетов сведены в таблицу:

i	$x_\imath$	$n_\imath$	$p_\imath$	$np_\imath$	$(n_\imath - np_i)^2$	$\dfrac{(n_\imath - np_\imath)^2}{np_\imath}$
1	< 75,84	9	0,125	12,5	12,25	1,00
2	75,84–86,18	16	0,125	12,5	12,25	1,00
3	86,18–99,37	24	0,125	12,5	132,25	10,58
4	99,37–100,77	0	0,125	12,5	156,25	12,50
5	100,77–102,17	4	0,125	12,5	72,25	5,78
6	102,17–103,74	2	0,125	12,5	110,25	8,82
7	103,74–105,85	6	0,125	12,5	42,25	3,38
8	> 105,85	39	0,125	12,5	702,25	56,18
		100	1,0			$\chi^2 = 99{,}24$

Со всей очевидностью критерий отклоняет гипотезу нормальности распределения, что видно по совершенно неестественному для нормального распределения правому „хвосту“ эмпирического распределения, делающего его явно несимметричным. Критическое значение статистики в этом случае при $f = k - 1 - 2 = 8 - 1 - 2 = 5$ равно $\chi^2_{0,9} = 5 \cdot \left(1 - \dfrac{2}{9 \cdot 5} + 1{,}28 \cdot \sqrt{\dfrac{2}{9 \cdot 5}}\right)^3 = 9{,}2$. Явное неравенство $\chi^2 = 99{,}24 > \chi^2_{0,9}(5) = 9{,}2$ отклоняет нулевую гипотезу.

Применение упрощенного критерия Романовского $R = \dfrac{22{,}9217 - 7}{\sqrt{2 \cdot 7}} = 4{,}25 > 3$ также отклоняет гипотезу.

3.1.1.2. Критерий числа пустых интервалов

Этот не очень мощный, но простой для применения критерий рассмотрен в [11]. Имеем выборку $x_1, \ldots, x_n$, разделенную на k интервалов в соответствии с гипотетическим распределением. Обозначим через a_0 число оставшихся пустыми интервалов ($n_\imath = 0$). Тогда имеет место

$$\mathbf{P}(a_0 = \gamma) = \frac{1}{k^n} C_k^\gamma \sum_{i=0}^{k-\gamma} C_{k-\gamma}^i (-1)^i (k - \gamma - 1)^n,$$

$$\text{где} \quad \gamma = c, c+1, \ldots, k-1; \quad c = \max(0, k-n).$$

Математическое ожидание и стандартное отклонение числа пустых интервалов равны соответственно

$$\mathbf{M}(a_0) = k\left(1 - \frac{1}{k}\right)^n;$$

$$\mathbf{D}(a_0) = k(k-1)\left(1 - \frac{2}{k}\right)^n + k(k-1)\left(1 - \frac{2}{k}\right)^n + k\left(1 - \frac{1}{k}\right)^n + k^2\left(1 - \frac{1}{k}\right)^{2n}.$$

При $k, n \to \infty$, $\rho = \dfrac{n}{k} > 0$ распределение a_0 асимптотически нормально с параметрами

$$\mathbf{M}(a_0) = ke^{-\rho}, \quad \mathbf{D}(a_0) = k\left[e^{-p} - e^{-2\rho}(1+\rho)\right].$$

Рекомендуется выбирать $\rho = 1{,}255$ или $k \approx 0{,}8n$ (наилучшая сходимость к нормальному распределению). Тогда можно утверждать, что величина

$$U = \frac{a_0 - ke^{-\rho}}{\left[k(e^{-\rho} - e^{-2\rho}(1+\rho))\right]^{\frac{1}{2}}}$$

имеет стандартное нормальное распределение и, следовательно, критическое значение статистики равно

$$a_0(\alpha) = 0{,}285k + u_\alpha 0{,}319\sqrt{k}.$$

Если эмпирическое значение числа пустых интервалов превысит $a_0(\alpha)$ (α — уровень достоверности), то с вероятностью α можно утверждать, что нулевая гипотеза отклоняется.

Возможна модификация этого критерия для проверки совпадения двух эмпирических распределений, т. е. для проверки гипотезы $H_0\colon F_n(x) = G_n(x)$. Рассмотрим выборку $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$, упорядоченную по возрастанию. На этой совокупности определим интервалы $I_1 \to (-\infty, x_1)$, $I_2 \to (x_1, x_2)$, $\ldots$, $I_{n+1}(x_n, \infty)$. Предположим, что имеется вторая выборка данных объема m, из которой r_i членов попадает в I_i-й интервал разбиения первой выборки. Обозначим через a_0' число оставшихся пустыми ячеек. Тогда

$$\mathbf{P}(a_0' = a) = \frac{C_{n+1}^a C_{m-1}^a}{C_{n+m}^n}, \quad \text{где} \quad a = c, c+1, \ldots, n; \quad c = \max(0, n+1-m).$$

В пределе

$$\mathbf{M}(a_0') = \frac{n+1}{n+l}; \quad \mathbf{D}(a_0') = \frac{(n+1)l^2}{(1+l)^3}; \quad l = \frac{m}{n}$$

$$\text{и} \quad U^* = \frac{a_0' - M(a_0')}{\sum D(a_0')} = \frac{a_0' - \dfrac{n+1}{n+l}}{\dfrac{l}{1+l}\sqrt{\dfrac{n+1}{n+l}}} \to N(0{,}1).$$

Следовательно, критическая величина $a_0'(\alpha)$ равна

$$a_0'(\alpha) = \frac{n+1}{n+l} + u_\alpha \frac{l}{1+l}\sqrt{\frac{n+1}{n+l}},$$

где u_α — α-квантиль стандартного нормального распределения.

Задача 92. *Имеются выборки данных:*

x_i:	1,	3,	5,	7,	9,	11,	13,	18,	21,	23,	25,	30,	32,	34,	35,	37
y_i:	−8,	−7,	4,	5,	6,	7,	8,	12,	14,	15,	18,	19,	20,	21,	23,	25

Необходимо проверить критерием числа пустых интервалов при достоверности $\alpha = 0{,}90$ *гипотезу о том, что случайная величина* x *подчиняется экспоненциальному распределению с параметром* $\dfrac{1}{\lambda} = 20$. *Проверить гипотезу о совпадении распределений случайных величин* x *и* y.

Имеем теоретическое распределение $F(x) = 1 - e^{-\frac{x}{20}}$.

Находим $k \approx 0{,}8 \cdot n = 0{,}8 \cdot 16 = 12$. Примем для упрощения расчетов $k = 10$ и $p_i = 0{,}1$ в каждом интервале. Тогда для границ интервалов имеем уравнение $-20 \cdot \ln[0{,}1 \cdot (i-1)] = x_i$.

Например, для x_2 имеем $x_2 = -20 \cdot \ln[1 - 0{,}1 \cdot (2-1)] = 2{,}1$.

Получаем 10 интервалов:

$$0 \div 2{,}1; \quad 2{,}1 \div 4{,}46; \quad 4{,}46 \div 7{,}13; \quad 7{,}13 \div 10{,}2; \quad 0{,}2 \div 13{,}86;$$
$$0{,}86 \div 18{,}3; \quad 18{,}3 \div 24{,}1; \quad 24{,}1 \div 32{,}19; \quad 32{,}19 \div 46{,}05; \quad 46{,}05 \div \infty.$$

Очевидно, что пустым остался один последний интервал, т. е. $a_0 = 1$.

Вычисляем

$$M(a_0) = 10 \cdot \left(1 - \frac{1}{10}\right)^{16} = 1{,}85;$$

$$D(a_0) = 10 \cdot 9 \cdot \left(1 - \frac{2}{10}\right)^{16} + 10 \cdot \left(1 - \frac{1}{10}\right) - 100 \cdot \left(1 - \frac{1}{10}\right)^{32} = 0{,}952.$$

Следовательно, $u = \dfrac{1 - 1{,}85}{\sqrt{0{,}952}} = -0{,}87$. Критическое значение статистики при $\alpha = 0{,}90$ ($u_{0,90} = 1{,}28$) равно

$$a_0(0{,}9) = 1{,}85 + 1{,}28 \cdot \sqrt{0{,}952} = 3{,}1.$$

Так как $a_0 = 1 < a_0(0{,}9) = 3{,}1$, нулевая гипотеза отклоняется.

Теперь разобьем элементы первой выборки на 16 интервалов и проверим попадание в них элементов второй выборки. Видим, что незаполненными остаются 7 интервалов, т. е. $a_0' = 7$. Тогда при $l = \dfrac{m}{n} = 1$ имеем

$$a_0'(0{,}9) = \frac{6+1}{1+1} + 1{,}28 \cdot \frac{1}{1+1} \cdot \sqrt{\frac{16+1}{1+1}} = 10{,}366.$$

Так как $a_0'(0{,}90) = 10{,}366 > a_0' = 7$, нулевая гипотеза не отклоняется.

3.1.1.3. Квартильный критерий Барнетта–Эйсена

В [208] предложен простой непараметрический критерий проверки согласия двух эмпирических распределений. Если разница в эмпирических распределениях является следствием разницы в параметрах положения и дисперсиях, то квартильный критерий может быть более эффективным, чем другие известные критерии.

Предположим, имеются две выборки $x_1, \ldots, x_m$ и $y_1, \ldots, y_n$ и $(m+n)$ делится на 4. Объединим обе выборки в одну объема $(m+n)$. Обозначим через a_i $(i = 1, 2, 3, 4)$ — число членов выборки x в i-й квартили объединенной выборки. Если обе выборки принадлежат одному распределению, то a_i имеет гипергеометрическое распределение и

$$\mathbf{P}(a_i = k) = \frac{C_m^k C_n^{Q-k}}{C_{m+n}^Q}; \quad \mathbf{M}(a_i) = \frac{m}{4}; \quad \mathbf{D}(a_i) = \frac{3mn}{16m+n-1}; \quad Q = \frac{m+n}{4}.$$

Введем обозначения

$$s_1 = a_1 + a_4; \quad d_0 = a_4 - a_1; \quad d_1 = a_3 - a_2.$$

Тогда имеют место соотношения:

$$\mathbf{M}(s) = \frac{m}{2}; \quad \mathbf{M}(d_0) = \mathbf{M}(d_1) = 0;$$

$$\mathbf{D}(s) = \frac{mn}{4(m+n-1)}; \quad \mathbf{D}(d_0) = \mathbf{D}(d_1) = \frac{mn}{2(m+n-1)}.$$

Введем обозначения

$$\tilde{s} = \frac{s - \dfrac{m}{2}}{\sqrt{D(s)}}; \quad \tilde{d}_0 = \frac{d_0}{\sqrt{D(d_0)}}; \quad \tilde{d}_1 = \frac{d_1}{\sqrt{D(d_1)}}.$$

В качестве критерия согласия предлагается статистика

$$D = (\tilde{s})^2 + (\tilde{d}_0)^2 + (\tilde{d}_1)^2,$$

имеющая при $m, n \to \infty$ распределение хи-квадрат с $f = 3$ степенями свободы. Если $D > \chi^2_\alpha(3)$, то гипотеза H_0 отклоняется.

Задача 93. *Имеются две выборки объема $m = n = 16$ каждая:*

x_i:	1,	3,	5,	7,	9,	11,	13,	18,	21,	23,	25,	30,	32,	34,	35,	37
y_i:	−8,	−7,	4,	5,	6,	7,	8,	12,	14,	15,	18,	19,	20,	21,	23,	25

Необходимо проверить гипотезу о согласии эмпирических распределений в выборках критерием Барнетта–Эйсена.

Составим объединенную выборку, отмечая в ней элементы обеих выборок с их порядковыми номерами

−8	−7	1	3	4	5	5	6	7	7	8	9	11	12	13	14
y_1	y_2	x_1	x_2	y_3	x_3	y_4	y_5	x_4	y_6	y_7	x_5	x_6	y_8	x_7	y_9

15	18	18	19	20	21	21	23	23	25	25	30	32	34	35	37
y_{10}	y_{11}	x_8	y_{12}	y_{13}	y_{14}	x_9	y_{15}	x_{10}	x_{11}	y_{16}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}

Имеем $Q = \dfrac{m+n}{4} = 8$. В первой квартили (первых 8 членах объединенной выборки) находятся $a_1 = 3$ элемента выборки x, во второй $a_2 = 4$; далее $a_3 = 2$ и $a_4 = 7$.

Имеем

$$s = a_1 + a_4 = 10; \quad d_0 = a_4 - a = 4; \quad d = a_3 - a = -2;$$

$$\mathbf{M}(s) = \frac{6}{2} = 8; \quad \mathbf{M}(d_0) = \mathbf{M}(d_1) = 0; \quad \mathbf{D}(s) = \frac{16 \cdot 16}{4 \cdot 31} = 2{,}0645 \ (\sqrt{\mathbf{D}(s)} = 1{,}4368);$$

$$\mathbf{D}(d_0) = \mathbf{D}(d_1) = \frac{16 \cdot 16}{2 \cdot 31} = 4{,}129 \ (\sqrt{\mathbf{D}(d_0)} = 2{,}032);$$

$$\tilde{s} = \frac{10-8}{1{,}4368} = 1{,}39198 \ (\tilde{s}^{\,2} = 1{,}9376); \quad \tilde{d}_0 = \frac{4}{2{,}032} = 1{,}9686 \ (\tilde{d}_0^2 = 3{,}875);$$

$$\tilde{d}_1 = -\frac{2}{2{,}032} = -0{,}9842 \ (\tilde{d}_1^2 = 0{,}9687); \quad \mathbf{D} = \tilde{s} + \tilde{d}_0^2 + \tilde{d}_1^2 = 6{,}7813.$$

Для $\alpha = 0{,}90$ и $f = 3$ имеем $\chi^2_{0,90}(3) = 6{,}251$.

Так как $D = 6{,}7813 > \chi^2_{0,90}(3)$, нулевая гипотеза отклоняется.

3.1.2. Критерии, основанные на сравнении теоретической и эмпирической функций распределения вероятностей

Обозначим через $F(x)$ $(x_i < x_{i+1})$ эмпирическую функцию распределения вероятностей, а через $\Phi(x)$ — теоретическую функцию распределения $\left(x_i = \Phi^{-1}\left(\frac{i-0{,}5}{n}\right)\right)$.

Расстояние между эмпирической и теоретической функциями распределения вероятностей является весьма эффективной статистикой для проверки гипотез о виде закона распределения вероятностей случайной величины.

Среди известных критериев согласия такого типа отметим серию критериев, использующих различные варианты анализа расстояния между $F(x)$ и $\Phi(x)$:

— критерий Джини (1941 г.)

$$\int |F(x) - \Phi(x)|\, dx;$$

— критерий Крамера–фон Мизеса (1928 г.)

$$\int \{F(x) - \Phi(x)\}^2\, dx;$$

— критерий Колмогорова–Смирнова (1933 г.)

$$\sup_{-\infty<x<\infty} |F(x) - \Phi(x)|;$$

— критерий Смирнова–Крамера–фон Мизеса (1936 г.)

$$\int \{F(x) - \Phi(x)\}^2 d\Phi(x);$$

— критерий Андерсона–Дарлинга (1952 г.)

$$\int \frac{\{F(x) - \Phi(x)\}^2}{\Phi(x)\{1 - \Phi(x)\}}\, d\Phi(x);$$

— критерий Купера (1960 г.)

$$\sup_{-\infty<x<\infty} \{F(x) - \Phi(x)\} + \sup_{-\infty<x<\infty} \{\Phi(x) - F(x)\};$$

— критерий Ватсона (1961 г.)

$$\int \left\{F(x) - \Phi(x) - \int [F(x) - \Phi(x)] d\Phi(x)\right\} d\Phi(x);$$

— критерий Фроцини (1978 г.)

$$\int |F(x) - \Phi(x)|\, d\Phi(x).$$

Из приведенного перечня видно, что исследователи весьма изобретательны, конструируя мыслимые и немыслимые варианты критериев. Однако не только из любви к собственно математической статистике они делают это. Разные критерии имеют различную мощность по отношению к различным альтернативам, т. е. предположениям, противостоящим выдвинутой (нулевой) гипотезе. Для того, чтобы наилучшим способом противостоять наиболее „опасной“ альтернативе, необходим достаточно широкий арсенал статистических инструментов. Проблема выдвижения альтернатив и оценка их „нежелательности“ не является задачей математической статистики, это проблема иных наук, занимающих исследователя.

Еще несколько важных замечаний, после чего приступим к изложению и иллюстрации основных практических критериев.

Если случайная величина имеет функцию распределения $\Phi(x)$, определенную с точностью до параметров, то случайная величина $\Phi(x_i)$ распределена равномерно на интервале $[0, 1]$. Таким образом, критерий согласия трансформируется в критерий проверки равномерности распределения случайной величины $\Phi(x_i)$ на интервале $[0, 1]$. В этом смысле все критерии согласия являются также критериями проверки равномерности распределения.

Следует избегать основной ошибки, совершаемой, к сожалению, подавляющей массой исследователей и инженеров. Общие критерии согласия, которые мы будем рассматривать далее, предполагают знание теоретического закона распределения с точностью до параметров. Так как это бывает редко в реальных ситуациях, исследователь мгновенно разрешает возникшую проблему простейшим способом — он проводит оценку параметров по самой выборке. Этого делать нельзя, так как достоверность полученных таким образом статистических выводов может быть сильно искажена [209]. Возможные варианты разрешения проблем, возникающих в случае отсутствия надежной информации о параметрах распределения, приведены в разделах, посвященных специальным критериям нормальности и экспоненциальности распределения. Поэтому, если вам необходимо проверить, нормально ли распределение полученных выборочных данных, а параметры гипотетического нормального распределения вы собираетесь также оценивать по выборке, не спешите воспользоваться общим критерием согласия из раздела 3.2.1. Пропустите этот раздел и обратитесь к разделу 3.2.2, в котором вы найдете квалифицированные рекомендации для такой ситуации.

3.1.2.1. Критерий Колмогорова–Смирнова

Пусть $F_n(x)$ — эмпирическая функция распределения случайной величины x, представленной выборкой $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$:

$$F_n(x) = \begin{cases} 0, & x < x_1; \\ \dfrac{i}{n}, & x_i \leqslant x < x_{i+1}, \quad 1 \leqslant i \leqslant n-1; \\ 1, & x \geqslant x_n. \end{cases}$$

Для проверки нулевой гипотезы $H_0\colon F_n(x) = \Phi(x)$, где $\Phi(x)$ — полностью определенная (с точностью до параметров) теоретическая функция распределения, рассматривается расстояние между эмпирической и теоретической функциями распределения

$$D_n = \sup_{|x|<\infty} |F_n(x) - \Phi(x)|; \quad D_n^+ = \sup_{|x|<\infty} (F_n(x) - \Phi(x));$$
$$D_n^- = -\inf_{|x|<\infty} (F_n(x) - \Phi(x)).$$

Здесь sup, inf — точные верхняя и нижняя границы соответствующих разностей.

Для практического применения используются формулы

$$D_n^+ = \max_{1\leqslant i\leqslant n}\left(\frac{i}{n} - \Phi(x)\right); \quad D_n^- = \max_{1\leqslant i\leqslant n}\left(\Phi(x) - \frac{i-1}{n}\right); \quad D_n = \max\left(D_n^+, D_n^-\right).$$

Колмогоров [210] нашел предельное распределение статистики $\sqrt{n}D_n$ (при $n \to \infty$). Смирнов [211] развил результаты Колмогорова на случай статистик D_n^+, D_n^-. Точные распределения статистик D_n^+, D_n^-, D_n приведены в [24, 25, 29, 57].

Между критическими значениями D_n и $D_n^+(D_n^-)$ существует соотношение $D_n^+(\alpha)(D_n^-(\alpha)) = D_n(2\alpha)$ (α — уровень значимости).

В качестве первого приближения можно использовать соотношение

$$D_n(\alpha) = \left\{\frac{1}{2}\ln\frac{2}{1-\alpha}\right\}^{-1}; \quad D_n^{+(-)}(\alpha) = \left\{\frac{1}{2}\ln\frac{1}{1-\alpha}\right\}^{-1}.$$

Если $D_n > D_n(\alpha)$, гипотеза согласия (H_0) отклоняется на уровне значимости α.

При $n \geqslant 20$ полезна аппроксимация [25]

$$\chi^2 = \frac{1}{9n}(6nD_n^{+(-)} + 1),$$

распределение которой удовлетворительно описывается распределением хи-квадрат с $f = 2$ степенями свободы.

При $n \geqslant 10$ необходимо использовать более точное приближение

$$D_n^{+(-)}(\alpha) = \left\{\frac{1}{2n}\left(y - \frac{2y^2 - 4y - 1}{18n}\right)\right\}^{\frac{1}{2}} - \frac{1}{6n} \approx \left(\frac{y}{2n}\right)^{\frac{1}{2}} - \frac{1}{6n},$$

где $y = -\ln\alpha$ для $D_n^{+(-)}$ и $y = -\ln(\alpha/2)$ для D_n, при $0{,}01 \leqslant \alpha \leqslant 0{,}2$ и $0{,}005 \leqslant \alpha$. Наиболее просты в приложениях результаты Стефенса [212], который предложил преобразования статистик $D_n^{+(-)}$, D_n, устраняющие зависимость их процентных точек $\tilde{D}_n$, $\tilde{D}_n^{+(-)}$ от объема выборки n:

$$\tilde{D}_n = D_n\left(\sqrt{n} + 0{,}275 - \frac{0{,}04}{\sqrt{n}}\right); \quad \tilde{D}_n = D_n\left(\sqrt{n} + 0{,}12 - \frac{0{,}11}{\sqrt{n}}\right);$$

$$\tilde{D}_n^{+(-)} = D_n^{+(-)}\left(\sqrt{n} + 0{,}12 + \frac{0{,}11}{\sqrt{n}}\right).$$

Первые две аппроксимации используются соответственно для нижних и верхних процентных точек. Критические значения статистик Стефенса приведены в табл. 53.

Таблица 53

Процентные точки статистик $\tilde{D}_n$ и $\tilde{D}_n^{+(-)}$ [215]

α	0,150	0,100	0,050	0,025	0,010
$\tilde{D}_n$	0,973	1,073	1,224	1,358	1,518
$\tilde{D}_n^{+(-)}$	1,138	1,224	1,358	1,480	1,628

Модифицированные статистики критерия Колмогорова–Смирнова, позволяющие применять их в некоторых частных случаях и для ситуаций с неизвестными параметрами гипотетических распределений, рассмотрены в [209, 213] и будут подробно проанализированы нами ниже в разделах, посвященных специальным критериям согласия. Ситуация, когда выборка усечена или цензурирована, рассмотрена в [214], в которой приведены и процентные точки критерия Колмогорова–Смирнова для разных степеней цензурирования.

Задача 94. *Проверить на уровне значимости* $\alpha = 0{,}10$ *нормальность распределения выборки* x_i: 4, 7, 8, 12, 18, 19, 21, 25, 30 *при условии что* $\Phi(x) = N(10; 5)$ (*т. е. гипотетическим распределением является нормальное распределение с параметрами* $\mu = 10$ *и* $\sigma = 5$).

Задача является демонстрационной — на практике критерий Колмогорова–Смирнова применяется при $n \geqslant 50$. Для вычисления значений функции нормального распреде-

ления $\Phi(x)$ можно использовать либо таблицы, либо аппроксимации (см. раздел 1.1.1). Необходимо выбирать аппроксимации повышенной точности, что позволит избежать накопления погрешностей аппроксимации. Результаты расчетов сведем в таблицу; напомним, что $\Phi(-z_i) = 1 - \Phi(z_i)$.

i	x_i	z_i	$\Phi(z_i)$	$\frac{i}{n}$	$\frac{i-1}{n}$	$\frac{i}{n} - \Phi(z_i)$	$\Phi(z_i) - \frac{i-1}{n}$
1	4	−1,20	0,1151	0,10	0,00	−0,0151	0,11510
2	7	−0,60	0,2743	0,12	0,10	−0,0743	0,17430
3	8	−0,40	0,3446	0,30	0,20	−0,0446	0,14460
4	9	−0,20	0,4207	0,40	0,30	−0,0207	0,12070
5	12	0,40	0,6554	0,50	0,40	−0,1554	0,25540
6	18	1,60	0,9452	0,60	0,50	−0,3452	0,44520
7	19	1,80	0,9641	0,70	0,60	−0,2641	0,36410
8	21	2,20	0,9866	0,80	0,70	−0,1866	0,18660
9	25	3,00	0,9986	0,90	0,80	−0,0986	0,19860
10	30	4,00	0,9996	1,00	0,90	0,00005	0,09996

Из таблицы следует, что

$$D_{10}^{+} = \max\left(\frac{i}{n} - \Phi(z_i)\right) = 0{,}0005; \quad D_{10}^{-} = \max\left(\Phi(z_i) - \frac{i-1}{n}\right) = 0{,}4452;$$

$$D_{10} = \max(D_{10}^{+}, D_{10}^{-}) = 0{,}4452.$$

Критическое значение равно $D_{10}(0{,}1) = \sqrt{\frac{1}{2\cdot 10}\cdot \ln\frac{2}{0{,}9}} = 0{,}1998.$

Так как $D_{10} = 0{,}4452 > D_{10}(\alpha) = 0{,}1998$, гипотеза нормальности отклоняется на уровне значимости $= 0{,}1$. Более точное приближение вычисляется по формуле $\chi^2 = \frac{(10 - 0{,}4453 + 1)^2}{9\cdot 10} = 8{,}536$.

Критическое значение $\chi^2(1-\alpha)$ при $f = 2$ степенях свободы равно 1,8856.

Так как $\chi^2 = 8{,}536 > \chi^2(1-\alpha) = \chi^2(0{,}90) = 1{,}8856$, гипотеза H_0 отклоняется. Рассмотрим более точную аппроксимацию

$$y = -\ln 0{,}1 = 2{,}302; \quad D_n^{-} = \sqrt{\frac{1}{20}\cdot\left(2{,}302 - \frac{2\cdot 2{,}302^2 - 4\cdot 2{,}302 - 1}{18\cdot 10}\right)} - \frac{1}{6\cdot 10} = 0{,}3224.$$

Так как $D_n^{-} = 0{,}4453 > D_n(\alpha) = 0{,}3224$, H_0 отклоняется. Далее находим статистику $\tilde{D}_n = 0{,}4453\left(\sqrt{10} + 0{,}12 - \frac{0{,}11}{\sqrt{n}}\right) = 1{,}477$. Ее критическое значение равно 1,224 (см. табл. 53 при $\alpha = 0{,}1$). Так как $\tilde{D}_n^{-} = 0{,}4453 > \tilde{D}_n^{-}(\alpha) = 1{,}224$, гипотеза H_0 отклоняется.

3.1.2.2. Критерий Смирнова–Крамера–фон Мизеса

Статистика критерия имеет вид [216–218]

$$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left\{F\left(x_i - \frac{2i-1}{2n}\right)\right\}^2,$$

где $F(x_i)$ — теоретическая функция распределения.

Необходимо помнить, что теоретическая функция распределения должна быть известна с точностью до параметров. Распространенная ошибка — использование в качестве $F(x)$ функции распределения с параметрами, оцениваемыми по выбор-

ке [209] — приводит к уменьшению величины критического значения статистики, т. е. к увеличению количества ошибок второго рода.

При объеме выборки $n > 40$ можно использовать приведенные в табл. 54 квантили распределения $n\omega^2$, которые следуют из его предельного распределения [218] (α — уровень значимости, принятый для проверки H_0).

Таблица 54

Квантили распределения $n\omega^2$ [218]

α	0,900	0,950	0,990	0,995	0,999
$n\omega^2(\alpha)$	0,3473	0,4614	0,7435	0,8694	1,1679

При $n < 40$ таблицей можно пользоваться [215, 218] с заменой $n\omega^2$ на $(n\omega^2)' = \left(n\omega^2 - \frac{0{,}4}{n} + \frac{0{,}6}{n^2}\right)\left(1 + \frac{1}{n}\right)$.

Известна аппроксимация распределения $n\omega_n^2$ с помощью χ^2-распределения [220] $n\omega_n^2 = a + b\chi^2$, где χ^2 — случайная величина, имеющая распределение хи-квадрат с f степенями свободы;

$$a = \frac{336n^2 - 959n + 609}{210(32n^2 - 61n + 30)}; \quad b = \frac{32n^2 - 61n + 30}{84n(4n - 3)}; \quad f = \frac{98}{5}\frac{n(4n-3)^3}{(32n^2 - 61n + 30)^2}.$$

Исследования авторов работы [221] позволяют сделать вывод о том, что на уровне значимости $\alpha > 0{,}01$ квантили точной и предельной функций распределения $n\omega^2$ практически неразличимы уже при объеме выборки $n \geqslant 4$. При таких уровнях значимости использование преобразования $\left(n\omega^2\right)'$ вместо $n\omega^2$ не дает существенных преимуществ.

Необходимо отметить, что и критерий $n\omega^2$ и критерий Колмогорова–Смирнова (см. раздел 3.1.2.1) подсчитываются по негруппированным выборкам (в отличие от критерия χ^2 — см. раздел 3.1.1.1). В [226] приведен анализ модификаций критериев типа $n\omega^2$ для группированных данных и показано, что группировка приводит к потере мощности критерия на 5–15%.

Задача 95. *В условиях задачи 94 проверить нулевую гипотезу нормальности распределения случайных величин критерием $n\omega^2$.*

Вычисления сводим в таблицу:

i	x_i	z_i	$F(z_i)$	$\frac{2i-1}{n}$	$F(z_i) - \frac{2i-1}{n}$	$\left\{F(z_i) - \frac{2i-1}{n}\right\}^2$
1	4	−1,2	0,1151	0,1	0,0151	$2{,}28 \cdot 10^{-4}$
2	7	−0,6	0,2743	0,3	−0,0257	$6{,}60 \cdot 10^{-4}$
3	8	−0,4	0,3446	0,5	−0,1554	0,02415
4	9	−0,2	0,4207	0,7	−0,2793	0,0780
5	12	0,4	0,6554	0,9	−0,2446	0,0598
6	18	1,6	0,9452	1,1	−0,1548	0,0240
7	19	1,8	0,9641	1,3	−0,3359	0,1183
8	21	2,2	0,9866	1,5	−0,5134	0,2636
9	25	3,0	0,9986	1,7	−0,7014	0,4919
10	30	4,0	0,9996	1,9	−0,9000	0,8100
						1,8706

Имеем $n\omega^2 = \dfrac{1}{12 \cdot 10} + 1{,}8706 = 1{,}8789$. При $\alpha = 0{,}9$ критическое значение равно $n\omega^2(0{,}9) = 0{,}3473$. Так как $n\omega^2(0{,}9) = 0{,}3473 < 1{,}8789$, гипотеза нормальности отклоняется. Вычислим более точный критерий

$$\left(n\omega^2\right)' = \left(1{,}8789 - \frac{0{,}4}{10} + \frac{0{,}6}{100}\right) \cdot \left(1 + \frac{1}{10}\right) = 2{,}029.$$

Видим, что результат тот же — H_0 отклоняется.

Теперь найдем χ^2-аппроксимацию критерия $n\omega^2$ и оценим ее точность. Вычисляем

$$a = \frac{336 \cdot 100 - 959 \cdot 10 + 609}{210 \cdot (32 \cdot 100 - 61 \cdot 10 + 30)} = 0{,}04474; \quad b = \frac{32 \cdot 100 - 61 \cdot 10 + 30}{84 \cdot 10 \cdot (4 \cdot 10 - 3)} = 0{,}0843;$$

$$f = \frac{98}{5} \cdot \frac{10 \cdot (4 \cdot 10 - 3)^3}{(32 \cdot 100 - 61 \cdot 10 + 30)^2} = 1{,}446; \quad n\omega^2(0{,}9) = 0{,}04474 + 0{,}0843 \cdot \chi^2_{0,9}(1{,}446).$$

Из таблиц имеем

$$\chi^2_{0,9}(1{,}446) \approx 3{,}65 \quad \text{и} \quad n\omega^2(0{,}9) = 0{,}04474 + 0{,}0843 \cdot 3{,}65 = 0{,}352.$$

Видим, что это значение близко к предельной квантили $n\omega^2(0{,}9) = 0{,}3473$, т. е. аппроксимация удовлетворительна.

Небольшое отступление: для удобства приведем квантили распределения χ^2 (табл. 55).

3.1.2.3. Критерий Реньи (*R*-критерий)

Отклонения на правом конце эмпирического распределения являются суммой многих отклонений, в том числе и расположенных на левом конце. Поэтому возможно, что эти отклонения будут принимать большие значения. Реньи [220] предложил критерий согласия, основанный на взвешивании статистики Колмогорова–Смирнова (см. раздел 3.1.2.1) обратным значением гипотетической теоретической функции распределения вероятностей $F(x)$. Статистики критерия Реньи имеют вид

$$R_n^+ = \sup_{F(x) \geqslant a} \frac{F_n(x) - F(x)}{F(x)} = \max_{F(x_i) \geqslant a} \frac{\dfrac{i}{n} - F(x_i)}{F(x_i)};$$

$$R_n^- = -\inf_{F(x) \geqslant a} \frac{F_n(x) - F(x)}{F(x)} = \max_{F(x_i) \geqslant a} \frac{F(x_i) - \dfrac{i-1}{n}}{F(x_i)};$$

$$R_n = \sup_{F(x) \geqslant a} \frac{|F_n(x) - F(x)|}{F(x)} = \max_{F(x_i) \geqslant a} (R_n^+, R_n^-), \quad (0 \leqslant a \leqslant 1).$$

Для больших n $(n \to \infty)$ [220]

$$\mathbf{P}\left\{\sqrt{\frac{na}{1-a}} R_n^+ < x\right\} = 2\Phi(x) - 1,$$

где $\Phi(x)$ — функция распределения $N(0,1)$ (см. раздел 1.1.1), т. е.

$$R_n^+(\alpha) = \sqrt{\frac{1-a}{na}}\, u_{1-\frac{\alpha}{2}}.$$

Таблица 55

Квантили распределения χ^2 (f — число степеней свободы) [25]

f	α							
	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99
1	0,0001	0,0010	0,0039	0,0158	2,706	3,841	5,024	6,6350
2	0,0201	0,0506	0,1030	0,2110	4,605	5,991	7,378	9,210
3	0,1150	0,2160	0,3520	0,5840	6,251	7,815	9,348	11,345
4	0,2970	0,4840	0,7110	1,0640	7,779	9,488	11,143	13,277
5	0,5540	0,8310	1,1450	1,6100	9,236	11,070	12,832	15,086
6	0,8720	1,2370	1,6350	2,2040	10,645	12,592	14,449	16,812
7	1,2390	1,6900	2,1670	2,8330	12,017	14,067	16,013	18,475
8	1,6460	2,1800	2,7330	3,4900	13,362	15,507	17,535	20,090
9	2,0880	2,7000	3,3250	4,1680	14,684	16,919	19,023	21,666
10	2,5580	3,2470	3,9400	4,8650	15,987	18,307	20,483	23,209
11	3,0530	3,8160	4,5750	5,5780	17,275	19,675	21,920	24,725
12	3,5710	4,4040	5,2260	6,3040	18,549	21,026	23,336	26,217
13	4,1070	5,0090	5,8920	7,0420	19,812	22,362	24,736	27,688
14	4,6600	5,6290	6,5710	7,7900	21,064	23,685	26,119	29,141
15	5,2290	6,2620	7,2610	8,5470	22,307	24,996	27,488	30,578
16	5,8120	6,9080	7,9620	9,3120	23,542	26,296	28,845	32,000
17	6,4080	7,5640	8,6720	10,0850	24,769	27,587	30,191	33,409
18	7,0150	8,2310	9,3900	10,8650	25,989	28,869	31,526	34,805
19	7,6330	8,9070	10,1170	11,6510	27,204	30,144	32,852	36,191
20	8,2600	9,5910	10,8510	12,4430	28,412	31,410	34,170	37,566
21	8,8970	10,2830	11,5910	13,2400	29,615	32,671	35,479	38,232
22	9,5420	10,9820	12,3380	14,0410	30,813	33,924	36,781	40,289
23	10,1960	11,6880	13,0910	14,8480	32,007	35,172	38,076	41,638
24	10,8560	12,4010	13,8480	15,6590	33,196	36,415	39,364	42,980
25	11,5240	13,1200	14,6110	16,4730	34,382	37,652	40,646	44,314
26	12,1980	13,8440	15,3790	17,2920	35,563	38,885	41,923	45,642
27	12,8790	14,5730	16,1510	18,1140	36,741	40,113	43,194	46,963
28	13,5650	15,3080	16,9280	18,9390	37,916	41,337	44,461	48,278
29	14,2560	16,0470	17,7080	19,7680	39,087	42,557	45,722	49,588
30	14,9530	16,7910	18,4930	20,5990	40,256	43,773	46,979	50,892

Отметим, что $R_n^+(\alpha) = R_n(2\alpha)$. При малых n имеет место соотношение [222]

$$\mathbf{P}(R_n^+ \geqslant x) = \frac{x}{1+x} \sum_{i=0}^{k} C_n^i \left(\frac{x + \frac{i}{n}}{1+x} \right)^{i-1} \left(\frac{1 - \frac{i}{n}}{1+x} \right)^{n-i},$$

где $x > 0$, $k = n - [na(1+x)] - 1$; $[\ldots]$ — целая часть числа.

При $a = 0$ имеем $\mathbf{P}(R_n^+ \geqslant x) = \dfrac{1}{1+x}$.

Задача 96. *В условиях задачи* 94 *проверить нулевую гипотезу нормальности распределения критерием Реньи.*

Вычисления сводим в таблицу.

Положим, что $a = 0{,}5$. Тогда из таблицы следует: $R_n^+ = 0{,}0005$; $R_n^- = 0{,}4710$ и $R_n = 0{,}4710$. Далее

$$\sqrt{\frac{na}{1-a}} \cdot R_n^- = \sqrt{\frac{10 \cdot 0{,}5}{0{,}5}} \cdot 0{,}4710 = 1{,}489;$$

i	x_i	z_i	$F(z_i)$	$\frac{i}{n}$	$\frac{i-1}{n}$	$\frac{i}{n} - F(z_i)$	$\frac{\frac{i}{n} - F(z_i)}{F(z_i)}$	$F(z_i) - \frac{i-1}{n}$	$\frac{F(z_i) - \frac{i-1}{n}}{F(z_i)}$
1	4	−1,2	0,11510	0,1	0,0	−0,0151	−0,1312	0,11510	1,00000
2	7	−0,6	0,27430	0,2	0,1	−0,0743	−0,2709	0,17430	0,63540
3	8	−0,4	0,34460	0,3	0,2	−0,0446	−0,1294	0,14460	0,41960
4	9	−0,2	0,42070	0,4	0,3	−0,0270	−0,0642	0,12070	0,28690
5	12	0,4	0,65540	0,5	0,4	−0,1554	−0,2371	0,25540	−0,38970
6	18	1,6	0,94520	0,6	0,5	−0,3452	−0,3652	0,44520	0,47100
7	19	1,8	0,96410	0,7	0,6	−0,2641	−0,2739	0,36410	0,37760
8	21	2,2	0,98660	0,8	0,7	−0,2866	−0,2905	0,28660	0,29050
9	25	3,0	0,99860	0,9	0,8	−0,0986	−0,0969	0,19860	0,19890
10	30	4,0	0,99996	1,0	0,9	0,0005	0,0005	0,09996	0,09996

$$\mathbf{P}\left(\sqrt{\frac{na}{1-a}} \cdot R_n^- < 1{,}4897\right) = 2(1{,}4897) - 1 = 0{,}864;$$

$$\mathbf{P}\left(\sqrt{\frac{na}{1-a}} \cdot R_n^- > 1{,}4897\right) = 1 - 0{,}864 = 0{,}136;$$

$$R_n^-(\alpha = 0{,}1) = \sqrt{\frac{1-a}{na}} \cdot u_{1-\frac{\alpha}{2}} = \frac{1}{\sqrt{10}} \cdot u_{0{,}95} = 0{,}521.$$

Так как $\sqrt{\frac{na}{1-a}} \cdot R_n^- = 1{,}489 > 0{,}521$, нулевая гипотеза отклоняется.

3.1.2.4. Критерий Андерсона–Дарлинга (критерий $n\Omega^2$)

По аналогии с критерием Реньи (см. раздел 3.1.2.3) Андерсон и Дарлинг [223] предложили критерий, использующий нормирование статистики критерия $n\omega^2$ (см. раздел 3.1.2.2) обратным значением теоретической функции распределения.

Статистика Андерсона–Дарлинга имеет вид

$$n\Omega^2 = -n - 2\sum_{i=1}^{n}\left\{\frac{2i-1}{2n}\ln F(x_i) + \left(1 - \frac{2i-1}{2n}\right)\ln[1 - F(x_i)]\right\} =$$
$$= -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left\{\ln F(x_i) + \ln\left[1 - F(x_{n-i+1})\right]\right\}.$$

Предельное распределение статистики $n\Omega^2$ (при $n \to \infty$) табулировано в [25, 218, 224]. В табл 56 приведены некоторые квантили предельного распределения $n\Omega^2$ (приближение приемлемо при $n > 50$).

Таблица 56

Квантили предельного распределения статистики $n\Omega^2$

α	0,90	0,95	0,975	0,99
$n\Omega^2(\alpha)$	1,94	2,50	3,08	3,88

Сходимость к предельному распределению становится лучше, если использовать вместо статистики $n\Omega^2$ ее модифицированную форму [226]

$$\left(n\Omega^2\right)' = \frac{n^2(n\Omega^2) + n + 1}{n^2 + n + 1}.$$

В [227] предложена модификация статистики Андерсона–Дарлинга в форме

$$\Omega^2 = n\Omega^2(U) + n\Omega^2(L),$$

где $n\Omega^2(U)$ — версия критерия $n\Omega^2$ для правого (верхнего) „хвоста“; $n\Omega^2(L)$ — версия критерия $n\Omega^2$ для левого (нижнего) „хвоста“.

Значения $n\Omega^2(U)$ и $n\Omega^2(L)$ вычисляются по формулам

$$n\Omega^2(U) = \frac{n}{2} - 2\sum_{i=1}^{n} F(x_i) - \sum_{i=1}^{n}\left[2 - \frac{2i-1}{n}\right]\ln\left[1 - F\left(x_i\right)\right];$$

$$n\Omega^2(L) = -\frac{3n}{2} + 2\sum_{i=1}^{n} F(x_i) - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\ln F(x_i).$$

Распределение величины $n\Omega^2(U)$ может быть вычислено по формуле [227]

$$n\Omega^2(U)_n = n\Omega^2(U)_\infty\left(1 + \frac{0{,}3}{\sqrt{n}}\right),$$

где $n\Omega^2(U)_\infty$ — предельное распределение, для которого справедлива аппроксимация

$$n\Omega_2(U)_\infty = 1 - \frac{1}{1 + \exp K(\alpha)},$$

где

$$K\left(\alpha\right) = 0{,}1170 - 0{,}03791y + 0{,}06318z + 0{,}09878yz + 0{,}009184y^2z - 0{,}0000742y^4z;$$

$$y = \ln\frac{\alpha}{1-\alpha}; \quad z = \left(1 + \frac{0{,}3}{\sqrt{n}}\right)^{-1}.$$

Значения $n\Omega^2(L)$ могут быть вычислены из условия симметрии. Модифицированный критерий Андерсона–Дарлинга более чувствителен к поведению функции распределения вероятностей на ее хвостах.

Задача 97. *В условиях задачи* 94 *проверить нулевую гипотезу нормальности распределения вероятностей критерием Андерсона–Дарлинга.*

Вычисления сведем в таблицу (используем обозначение $\# = (2i-1)\{\ln F(z_i) + \ln[1 - \ln F(z_{n-i+1})]\}$):

i	x_i	z_i	$F(z_i)$	x_{n-i+1}	z_{n-i+1}	$F(z_{n-i+1})$	$\ln F(z_i)$	$\ln[1 - F(z_{n-i+1})]$	#
1	4	−1,2	0,1151	30	4,0	0,9999	−2,1619	−10,1266	−12,2885
2	7	−0,6	0,2743	25	3,0	0,9986	−1,2935	−6,5713	−27,9716
3	8	−0,4	0,3446	21	2,2	0,9866	−1,0653	−4,3125	−26,8890
4	9	−0,2	0,4207	19	1,8	0,9641	−0,0658	−3,3210	−29,9394
5	12	0,4	0,6554	18	1,6	0,9452	−0,4225	−2,9041	−29,9394
6	18	1,6	0,9452	12	0,4	0,6554	−0,0563	−1,0653	−12,3376
7	19	1,8	0,9641	9	−0,2	0,4207	−0,0366	−0,5459	−7,5725
8	21	2,2	0,9866	8	−0,4	0,3446	−0,0135	−0,4225	−6,5400
9	25	3,0	0,9986	7	−0,6	0,2743	−0,0014	−0,3206	−5,4740
10	30	4,0	0,9999	4	−1,2	0,1151	$-4 \cdot 10^{-5}$	−0,1223	−2,3245
									−161,2765

На основании данных таблицы имеем

$$n\Omega^2 = -10 - \frac{1}{10}\cdot(-161{,}2765) = 6{,}127; \quad (n\Omega^2)' = \frac{6{,}127\cdot 100 + 10 + 1}{100 + 10 + 1} = 5{,}619.$$

Из табл. 56 имеем $n\Omega^2(0{,}9) = 1{,}94$.

Так как $(n\Omega^2)' = 5{,}619 > n\Omega^2(0{,}9) = 1{,}94$, нулевая гипотеза нормальности распределения H_0 отклоняется.

3.1.2.5. Критерий Ватсона

Так как $y_i = F(x_i)$ имеет равномерное распределение на единичном интервале $[0,1]$, то проверка гипотезы равномерности распределения y_i равносильна проверке нулевой гипотезы о подчинении выборочных данных распределению $F(x)$. Поэтому предложенный Ватсоном [225] критерий для проверки таких гипотез в литературе чаще всего представлен как критерий равномерности распределения. Статистика критерия Ватсона имеет вид

$$U_n^2 = n\int\limits_{-\infty}^{\infty}\left\{F_n(x) - F(x) - \int\limits_{-\infty}^{\infty}[F_n(y) - F(y)]\,dF(y)\right\}^2 dF(x),$$

или в форме, удобной для расчетов:

$$U_n^2 = \sum_{i=1}^{n}\left(y_i - \frac{i-\frac{1}{2}}{n}\right)^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{2}\right)^2 + \frac{1}{12n}.$$

Необходимо помнить, что $y_i = F(x_i)$.

Приведем эквивалентную форму

$$U_n^2 = \sum_{i=1}^{n}\left(y_i - \frac{2i-1}{2n} - \bar{y} + \frac{1}{2}\right)^2 + \frac{1}{12n}, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Между критерием $n\omega^2$ Смирнова–Крамера–фон Мизеса (см. раздел 3.1.2.2) и критерием Ватсона существует простое соотношение

$$U_n^2 = n\omega^2 - n\left(\bar{y} - \frac{1}{2}\right)^2.$$

Для практических расчетов рекомендуется формула

$$U_n^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 - \frac{2}{n}\sum_{i=1}^{n} i y_i + (n+1)\bar{y} + \frac{n}{12}.$$

Статистика

$$(U_n^2)' = \left(U_n^2 - \frac{0{,}1}{n} + \frac{0{,}1}{n^2}\right)\left(1 + \frac{0{,}8}{n}\right)$$

уже при $n \geqslant 10$ имеет предельное распределение, критические нижние точки которого приведены в табл. 57 (α — уровень значимости).

Таблица 57

Нижние критические точки статистики Ватсона U_n^2 [228]

α	0,10	0,05	0,025	0,01
$(U_n^2)'$	0,152	0,187	0,221	0,267

Нулевая гипотеза отклоняется, если $(U_n^2)' < (U_n^2(\alpha))'$.

Процентные точки и полную функцию распределения U_n^2 можно найти в [218]. Там же, как и для распределения $n\omega^2$ (см. раздел 3.1.2.2), для U_n^2 предложена аппроксимация с помощью χ^2-распределения [219]

$$U_n^2 = a + b\chi^2(f),$$

где

$$a = \frac{21n - 56}{840\left(n - \dfrac{3}{2}\right)}; \quad b = \frac{1}{42n}\left(n - \frac{3}{2}\right); \quad f = \frac{49n(n-1)}{20\left(n - \dfrac{3}{2}\right)^2}.$$

Задача 98. *Проверить гипотезу нормальности распределения вероятностей в условиях задачи* 94 *критерием Ватсона.*

Имеем

$$y_{1f} = F(x_1) = F\left(\frac{4-10}{5}\right) = F(-1{,}2) = 0{,}1151;$$

$$y_2 = F(x_2) = F(-0{,}6) = 0{,}2743; \qquad y_3 = F(x_3) = F(-0{,}4) = 0{,}3446;$$
$$y_4 = F(x_4) = F(-0{,}2) = 0{,}4207; \qquad y_5 = F(_5) = F(0{,}4) = 0{,}6554;$$
$$y_6 = F(x_6) = F(1{,}6) = 0{,}9452; \qquad y_7 = F(x_7) = F(1{,}8) = 0{,}9641;$$
$$y_8 = F(x_8) = F(2{,}2) = 0{,}9866; \qquad y_9 = F(x_9) = F(3{,}0) = 0{,}9986;$$
$$y_{10} = F(x_{10}) = F(4{,}0) = 0{,}99996.$$

Тогда $\bar{y} = 0{,}670456$ и

$$U_n^2 = \sum_{\imath=1}^{n}\left(y_\imath - \frac{2i-1}{n} - 0{,}670456 + 0{,}5\right)^2 + \frac{1}{120} = 0{,}1290.$$

Вычисляем модифицированную форму критерия

$$\left(U_n^2\right)' = \left(0{,}129 - \frac{0{,}1}{10} + \frac{0{,}1}{100}\right) \cdot \left(1 + \frac{0{,}8}{10}\right) = 0{,}1296.$$

Из табл. 57 при $\alpha = 0{,}1$ находим $U_n^2(0{,}1) = 0{,}152$.

Так как $(U_n^2)' = 0{,}1296 < (U_n^2(0{,}1))' = 0{,}152$, гипотеза нормальности распределения случайных величин отклоняется.

Воспользуемся теперь χ^2-аппроксимацией:

$$a = \frac{21 \cdot 10 - 56}{840 \cdot \left(10 - \dfrac{3}{2}\right)} = 0{,}02157; \quad b = \frac{1}{42 \cdot 10} \cdot \left(10 - \frac{3}{2}\right) = 0{,}02023;$$

$$f = \frac{49}{20} \cdot \frac{10 \cdot 9}{\left(10 - \dfrac{3}{2}\right)^2} \approx 3.$$

Из табл. 55 находим $\chi^2_{0,9}(3) = 6{,}251$ и вычисляем

$$U_n^2(0{,}1) = 0{,}02157 + 0{,}02029 \cdot 6{,}251 = 0{,}148,$$

что очень близко к табличной величине $U_n^2(0{,}1) = 0{,}152$.

3.1.2.6. Критерий Купера

Купер [229] предложил расширенную статистику критерия типа Колмогорова–Смирнова (см. раздел 3.1.2.1)

$$V_n = \sup_{-\infty<x<\infty}\{F_n(x) - F(x)\} - \inf_{-\infty<x<\infty}\{F_n(x) - F(x)\}, \quad \text{или} \quad V_n = D_n^+ D_n^-,$$

где

$$D_n^+ = \max\left\{\frac{i}{n} - F(x)\right\} \quad \text{и} \quad D_n^- = \max\left\{F(x) - \frac{i-1}{n}\right\}$$

— статистики Колмогорова–Смирнова.

Верхние квантили предельного распределения V_n приведены в [230].

Если $A = \max\left(\frac{i}{n} - y_i\right)$, $B = \max\left(y_i - \frac{i-1}{n}\right)$, где $y_i = F(x_i)$, то расчетная формула может быть записана в простейшей форме

$$V_n = A + B.$$

К предельному распределению быстро сходится модифицированная форма статистики критерия

— для верхних процентных точек $V_n' = \left(\sqrt{n} + 0{,}155 + \frac{0.24}{\sqrt{n}}\right)V_n$;

— для нижних процентных точек $V_n' = \left(\sqrt{n} + 0{,}41 - \frac{0{,}26}{\sqrt{n}}\right)V_n$.

Критические значения для верхних процентных точек приведены в табл. 58.

Таблица 58

Критические значения $V_n'(\alpha)$ для верхних процентных точек статистики Купера [212]

α	0,10	0,05	0,025	0,01
$V_n'(\alpha)$	1,62	1,747	1,862	2.001

Если $V_n' < V_n'(\alpha)$, то нулевая гипотеза принимается. Подробно применение критерия Купера и расширенные таблицы его квантилей приведены в [231]. Применение критерия Купера в частном случае проверки соответствия эмпирического распределения распределению Вейбулла рассмотрено в [212].

Задача 99. *В условиях задачи* 94 *проверить гипотезу нормальности распределения критерием Купера.*

При решении задачи 94 было получено $D_n^+ = 0{,}00005$ и $D_n^- = 0{,}4452$, откуда $V_n = D_n^+ + D_n^- = 0{,}44525$, или в модифицированной форме

$$V_n' = 0{,}44525 \cdot \left(\sqrt{10} + 0{,}155 - \frac{0{,}24}{\sqrt{10}}\right) = 1{,}443.$$

Из табл. 58 имеем $V_n'(0{,}1) = 1{,}62$. Так как $V_n' = 1{,}443 < V_n'(0{,}1) = 1{,}62$, нулевая гипотеза нормальности исходного распределения случайных величин не отклоняется (делать из этого вывода трагедию не стоит — критерий Купера хорошо работает только при $n \geqslant 20$).

3.1.2.7. Критерий согласия Дарбина

Наиболее известный критерий согласия — критерий χ^2 (см. раздел 3.1.1.1) гибок, легко используется, но имеет элемент произвола в выборе границ группирования экспериментальных данных. Критерий Колмогорова–Смирнова (см. раздел 3.1.2.1) [232] свободен от этих недостатков и имеет хорошую асимптотическую мощность по сравнению с альтернативами, определенными в терминах расстояния между функциями распределения. Однако исследования показывают [233], что на практике для выборок среднего объема он часто непригоден, в отличие от критерия χ^2.

Критерии типа Колмогорова–Смирнова хороши, когда альтернативное распределение таково, что разница между ним и исходным (например, разница в средних) велика. Однако если разница между средними и дисперсиями невелика, но две частотные функции заметно отличаются формой, то критерий Колмогорова–Смирнова не будет мощным критерием.

В [233] предлагаются новые критерии, свободные от распределения, более мощные, чем критерий Колмогорова–Смирнова.

Пусть $F(x)$ — гипотетическая теоретическая функция распределения вероятностей, определенная с точностью до параметров. Обозначим $U_j = F(x_j)$, $j = 1, \ldots, n$. При справедливости гипотезы H_0 величина U_j должна быть распределена равномерно на единичном интервале $[0, 1]$.

Пусть $U_1 \leqslant U_2 \leqslant \ldots \leqslant U_n$ — порядковые статистики ряда U_j. Сформируем новую последовательность

$$C_1 = U; \quad C_j = U_j - U_{j-1},\ (j = 2, \ldots, n); \quad C_{n+1} = 1 - U_n;$$
$$g_j = (n + 2 - j)(C_{(j)} - C_{(j-1)}); \quad C_0 = 0, \quad j = 1, \ldots, n + 1,$$

где $C_{(j)}$ — порядковая статистика ряда C_j, т. е. j-е по величине значение C_j в упорядоченном по возрастанию ряду значений C_j.

Введем переменную $\omega_r = \sum\limits_{j=1}^{n+1} g_j$. Очевидно, что

$$\sum_{j=1}^{n+1} g_j = \sum_{j=1}^{n+1} C_{(j)} = 1; \quad \omega_j = C_{(1)} + \ldots + C_{(j-1)} + (n + 2 - j) \cdot C_{(j)}, j = 1, \ldots, n;$$
$$C_{(1)} \leqslant C_{(2)} \leqslant \ldots \leqslant C_{(n+1)}.$$

Сутью предлагаемых критериев является проверка равномерности распределения ω_j на интервале $[0, 1]$.

Рассмотрим различные варианты критериев проверки равномерности распределения ω_j, предложенные в [233].

3.1.2.7.1. Модифицированный медианный критерий

Статистика критерия

$$M_r = \frac{r}{n + 1 - r} \frac{1 - \omega_r}{\omega_r},$$

где $r = \dfrac{n+1}{2}$ (n — нечетное) и $r = \dfrac{n}{2}$ (n — четное).

При справедливости нулевой гипотезы статистика M_r распределена как $F_{2(n+1-r),2r}$, т. е. как случайная величина, имеющая F-распределение Фишера с $f_1 = 2(n + 1 - r)$ и $f_2 = 2r$ степенями свободы.

Гипотеза H_0 отклоняется с достоверностью α, если $M_r > F(\alpha)$, где $F(\alpha)$ — критическое значение F-распределения.

3.1.2.7.2. Модифицированный критерий Колмогорова–Смирнова

Статистика критерия

$$K_m = \max_{r=1,\ldots,n} \left(\frac{r}{n} - \omega_r \right).$$

Нулевая гипотеза отклоняется, если K_m превышает критическое значение статистики Колмогорова–Смирнова (см. раздел 3.1.2.2).

3.1.2.7.3. Модифицированный вероятностный критерий

В [234] показано, что $\prod_{j=1}^{n} U_j \to \exp\left(-\frac{1}{2}\chi^2\right)$, где χ^2 — хи-квадрат-распределенная случайная величина с $f = 2n$ степенями свободы (см. раздел 1.1.8). Тогда случайная величина

$$p_m = -2\ln\prod_{j=1}^{n}\omega_j$$

имеет χ^2-распределение с $f = 2n$ степенями свободы.

Если $p_m > \chi^2(2n)$, то гипотеза H_0 отклоняется с достоверностью α.

Мощность этого критерия исследована в [235], где показано, что этот критерий может иметь большую мощность для широкого класса альтернатив. Однако следует помнить, что статистика p_m сильно зависит от ошибок округления величин $U_1, \ldots, U_n$ и расчеты следует вести с максимально возможной точностью.

Сравнительный анализ рассмотренных критериев показывает, что:

— критерий Колмогорова–Смирнова уступает критерию χ^2;

— критерий M_r лучше критерия Колмогорова–Смирнова, но хуже критерия χ^2;

— критерий K_m мощнее, чем критерий Колмогорова–Смирнова, и не уступает критерию χ^2.

Задача 100. *В условиях задачи 94 проверить гипотезу нормальности распределения случайной величины критерием Дарбина.*

Имеем

$$U_1 = \Phi(-1{,}2) = 0{,}1151;\quad U_6 = \Phi(1{,}6) = 0{,}9452;$$
$$U_2 = \Phi(-0{,}6) = 0{,}2743;\quad U_7 = \Phi(1{,}8) = 0{,}9641;$$
$$U_3 = \Phi(-0{,}4) = 0{,}3446;\quad U_8 = \Phi(2{,}2) = 0{,}9866;$$
$$U_4 = \Phi(-0{,}2) = 0{,}4207;\quad U_9 = \Phi(3{,}0) = 0{,}9986;$$
$$U_5 = \Phi(0{,}4) = 0{,}6554;\quad U_{10} = (4{,}0) = 0{,}99996.$$

Сформируем новую последовательность

$$C_1 = U_1 = 0{,}1151;\quad C_2 = U_2 - U_1 = 0{,}1592;\quad C_7 = U_7 - U_6 = 0{,}0189;$$
$$C_3 = U_3 - U_2 = 0{,}0703;\quad C_8 = U_8 - U_7 = 0{,}0255;\quad C_4 = U_4 - U_3 = 0{,}0761;$$
$$C_9 = U_9 - U_8 = 0{,}0120;\quad C_5 = U_5 - U_4 = 0{,}2347;\quad C_{10} = U_{10} - U_9 = 0{,}00136;$$
$$C_6 = U_6 - U_5 = 0{,}2898;\quad C_{11} = 1 - U_{10} = 0{,}00004.$$

Ранжируем ряд C_j по возрастанию величины:

$$C_{(1)} = 0{,}00004;\quad C_{(2)} = 0{,}00136;\quad C_{(3)} = 0{,}0120;\quad C_{(4)} = 0{,}0189;$$
$$C_{(5)} = 0{,}0225;\quad C_{(6)} = 0{,}0703;\quad C_{(7)} = 0{,}0761;\quad C_{(8)} = 0{,}1151;$$
$$C_{(9)} = 0{,}1592;\quad C_{(10)} = 0{,}2347;\quad C_{(11)} = 0{,}2898.$$

Формируем ряд значений g_j:

$$g_1 = (n + 2 - 1)\cdot(C_{(1)} - 0) = 0{,}00044;$$
$$g_2 = 10\cdot(C_{(2)} - C_{(1)}) = 0{,}0132;\quad g_3 = 9\cdot(C_{(3)} - C_{(2)}) = 0{,}09576;$$
$$g_4 = 8\cdot(C_{(4)} - C_{(3)}) = 0{,}0552;\quad g_5 = 7\cdot(C_{(5)} - C_{(4)}) = 0{,}0252;$$
$$g_6 = 6\cdot(C_{(6)} - C_{(5)}) = 0{,}2868;\quad g_7 = 5\cdot(C_{(7)} - C_{(6)}) = 0{,}0290;$$
$$g_8 = 4\cdot(C_{(8)} - C_{(7)}) = 0{,}1560;\quad g_9 = 3\cdot(C_{(9)} - C_{(8)}) = 0{,}1323;$$
$$g_{10} = 2\cdot(C_{(10)} - C_{(9)}) = 0{,}1510;\quad g_{11} = 1\cdot(C_{(11)} - C_{(10)}) = 0{,}0591.$$

Окончательно имеем ряд значений ω_r:

$$\omega_1 = g = 0{,}00044;$$

$$\omega_2 = 0{,}00044 + 0{,}09576 = 0{,}01324; \quad \omega_3 = 0{,}01324 + 0{,}09576 = 0{,}10890;$$
$$\omega_4 = 0{,}1090 + 0{,}0552 = 0{,}1652; \quad \omega_5 = 0{,}1642 + 0{,}0252 = 0{,}1894;$$
$$\omega_6 = 0{,}1894 + 0{,}2868 = 0{,}4762; \quad \omega_7 = 0{,}4762 + 0{,}0290 = 0{,}5052;$$
$$\omega_8 = 0{,}5052 + 0{,}1560 = 0{,}6612; \quad \omega_9 = 0{,}6612 + 0{,}1323 = 0{,}7935;$$
$$\omega_{10} = 0{,}7935 + 0{,}1510 = 0{,}9445; \quad \omega_1 = 0{,}9445 + 0{,}0551 = 0{,}9996.$$

Модифицированный медианный критерий

Имеем $r = \dfrac{10}{2} = 5$, $M = \dfrac{5}{10+1-5} \cdot \dfrac{1-\omega_5}{\omega_5} = \dfrac{5}{6} \cdot \dfrac{1-0{,}1894}{0{,}1894} = 3{,}5665.$

Для степеней свободы $f_1 = 2 \cdot (n+1-r) = 12$ и $f_2 = 10$ ($\alpha = 0{,}90$) имеем из таблиц F-распределения (или аппроксимаций из раздела 1.1.10) $F_{12,10}(0{,}9) = 2{,}284$.

Так как $M_5 = 3{,}5665 > F_{12,10}(0{,}9) = 2{,}284$, нулевая гипотеза отклоняется.

Модифицированный критерий Колмогорова–Смирнова

Имеем разности

$$\frac{1}{n} - \omega_1 = 0{,}09956; \quad \frac{2}{n} - \omega_2 = 0{,}1868; \quad \frac{3}{n} - \omega_3 = 0{,}1910; \quad \frac{4}{n} - \omega_4 = 0{,}2358;$$
$$\frac{5}{n} - \omega_5 = 0{,}3106; \quad \frac{6}{n} - \omega_6 = 0{,}1238; \quad \frac{7}{n} - \omega_7 = 0{,}1948; \quad \frac{8}{n} - \omega_8 = 0{,}1388;$$
$$\frac{9}{n} - \omega_9 = 0{,}1065; \quad \frac{10}{n} - \omega_{10} = 0{,}0555.$$

Отсюда $K_m = 0{,}3106$ и из раздела 3.1.2.1 находим критическое значение статистики Колмогорова–Смирнова (при $\alpha = 0{,}10$). Оно равно для модифицированной статистики $\tilde{D}_n(0{,}1) = 1{,}073$. Тогда имеем

$$\tilde{K}_m = K_m \left(\sqrt{n} + 0{,}275 - \frac{0{,}04}{\sqrt{n}} \right) = 1{,}0715.$$

Значение $\tilde{K}_m = 1{,}0715$ очень близко к критическому $\tilde{K}_m(0{,}1) = 1{,}073$, поэтому гипотезу нормальности следует отклонить.

Модифицированный вероятностный критерий

$$p_m = -2 \ln \prod_{j=1}^{n+1} \omega_j = 39{,}633.$$

Из табл. 55 для $f = 2n = 20$ имеем $\chi^2_{20}(0{,}9) = 28{,}412$ ($\alpha = 0{,}90$, т. е. берется верхняя 10%-я точка распределения).

Так как $p_m = 39{,}693 > \chi^2_{20}(0{,}9) = 28{,}412$, нулевая гипотеза отклоняется.

3.1.2.8. Двухвыборочные критерии согласия

3.1.2.8.1. Двухвыборочный критерий Колмогорова–Смирнова

Рассматриваются выборки случайных величин

$$x\colon x_1, x_2, \ldots, x_{n_1} \quad \text{и} \quad y\colon y_1, y_2, \ldots, y_{n_2}.$$

Перед исследователем стоит вопрос: обе выборки извлечены из совокупности с одним и тем же законом распределения вероятностей? Говоря языком математической статистики, ему необходимо проверить нулевую гипотезу $H_0\colon F_{n_1}(x) = F_{n_2}(y)$ о совпадении функций распределения вероятностей в двух выборках. Статистики

критерия имеют вид [211]

$$D_n = \max|F_{n_1}(x) - F_{n_2}(y)|; \quad D_n^* = \max[F_{n_1}(x) - F_{n_2}(y)]$$

и определяются разностью двух эмпирических функций распределения вероятностей.

Распределение статистик D_n и D_n^* приведено в [25, 29]. Предельное распределение ($n_1, n_2 \to \infty$) табулировано в [29]. Критерием рекомендуется пользоваться для выборок объема $n \geqslant 100$. В этом случае справедливо соотношение

$$\mathbf{P}\left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \max[F_{n_1}(x) - F_{n_2}(y)] < z\right) \approx 1 - e^{-2z^2}.$$

Задача 101. *В результате сравнения двух выборок случайных величин X и Y, объемом $n_1 = 100$ и $n_2 = 300$ соответственно, получено максимальное расхождение их функций распределения вероятностей, равное 0,252. На уровне значимости $\alpha = 0{,}1$ проверить гипотезу о неразличимости функций распределения вероятностей в обеих выборках.*

Находим вероятность

$$\mathbf{P}\left(\sqrt{\frac{100 \cdot 300}{100 + 300}} \cdot \max\left[F_{100}(x) - F_{300}(y)\right] < 0{,}252\right) = 1 - \exp\left(-2 \cdot 0{,}252^2\right) = 0{,}119.$$

Так как полученная величина 0,119 превышает $\alpha = 0{,}1$, нулевая гипотеза отклоняется.

3.1.2.8.2. Критерий Катценбайссера–Хакля

В [236] предложен новый критерий, основанный на сравнении эмпирических функций распределения вероятностей, более мощный, чем критерий Колмогорова–Смирнова. Статистика критерия (T) определяется числом точек, в которых эмпирические функции распределения совпадают. Рассмотрим две выборки случайных величин равного объема x: $x_1, x_2, \ldots, x_n$ и y: $y_1, y_2, \ldots, y_n$.

Составляем из них новую, упорядоченную по возрастанию, выборку $\{z_i\}$ объема $N = 2n$. Определим сравниваемые эмпирические функции распределения следующим образом: $F_n(z_i) = \dfrac{n_x(i)}{n}$ $\left(G_n(z_i) = \dfrac{n_y(i)}{n}\right)$, где $n_x(i)$ $(n_y(i))$ — количество x-ов (y-ов), меньших или равных z_i.

В [236] показано, что

$$\mathbf{P}(T > t) = 2^t \frac{C_{2n-t}^{n-t}}{C_{2n}^n}, \quad t = 1{,}2, \ldots, n.$$

Математическое ожидание и дисперсия числа совпадающих точек равны

$$\mathbf{M}(T) = 2^{2n} \frac{1}{C_{2n}^n}; \quad \mathbf{D}(T) = 4n + 2 - 2^{2n} \frac{1}{C_{2n}^n} \left(1 + 2^{2n} \frac{1}{C_{2n}^n}\right).$$

Для больших выборок ($n \to \infty$)

$$\mathbf{M}(T) = \sqrt{\pi} n^{\frac{1}{2}} + \frac{1}{8}\sqrt{\pi} n^{-\frac{1}{2}} + \frac{1}{128}\sqrt{\pi} n^{-\frac{3}{2}};$$

$$\mathbf{D}(T) = (4 - \pi)\, n - \sqrt{\pi} n^{\frac{1}{2}} + \left(2 - \frac{\pi}{4}\right) - \frac{\sqrt{\pi}}{8} n^{-\frac{1}{2}} - \frac{1}{32}\pi n^{-1};$$

$$\mathbf{P}(T > t) = \sqrt{\frac{n}{M}} \left[1 - \frac{1}{96M}\left(x^4 - 12x^2 + 12\right)\right] \left(1 + \frac{1}{8n} + \frac{1}{128n^2}\right) e^{-\frac{x^2}{4}},$$

где $M = \dfrac{2n - t}{2}$ и $x = \dfrac{t}{\sqrt{M}}$.

Приближение удовлетворительно, когда $n \geqslant 5$. Если для полученного значения T вероятность $\mathbf{P}(T \leqslant t)$ будет меньше уровня значимости α, то нулевая гипотеза отклоняется и функции распределения не признаются совпадающими.

Задача 102. *Даны две выборки случайных величин:*

x_i: 10 12 14 14 18 21 27 28 35 36

y_i: 1 8 10 14 26 27 28 30 31 40

Проверить совпадение законов распределения вероятностей двухвыборочным критерием Катценбайссера–Хакля на уровне значимости $\alpha = 0{,}1$.

Имеем ряд:

$z_{(i)}$: 1(y), 8(y), 10(x), 10(y), 12(x), 14(x), 14(x), 14(y), 18(x), 21(x), 26(y), 27(x), 27(y), 28(x), 28(y), 30(y), 31(y), 35(x), 36(x), 40(y).

В скобках указана принадлежность значения первоначальным выборкам x и y. Вычисляем выборочное распределение. Результаты вычислений приведены в таблице:

i	$F_n(z_{(i)})$	$G_n(z_{(i)})$	i	$F_n(z_{(i)})$	$G_n(z_{(i)})$	i	$F_n(z_{(i)})$	$G_n(z_{(i)})$
1	0,0	0,1	8	0,4	0,4	15	0,8	0,7
2	0,0	0,2	9	0,5	0,4	16	0,8	0,8
3	0,1	0,3	10	0,6	0,4	17	0,8	0,8
4	0,1	0,3	11	0,6	0,5	18	0,9	0,9
5	0,2	0,3	12	0,7	0,6	19	1,0	0,9
6	0,4	0,4	13	0,7	0,6	20	1,0	1,0
7	0,4	0,4	14	0,8	0,7			

Из табл. видим, что функции распределения вероятностей совпадают в точках 0,4 ($i = 6, 7, 8$); 0,8 ($i = 16$); 0,9 ($i = 18$) и 1,0 ($i = 20$), т. е. $T = 5$.

Вычисляем

$$\mathbf{M}(T) = \sqrt{\pi n} + \frac{1}{8} \cdot \sqrt{\frac{\pi}{n}} + \frac{1}{128n} \cdot \sqrt{\frac{\pi}{n}} = 5{,}605 + 0{,}0700 + 4{,}379 \cdot 10^{-4} = 5{,}675;$$

$$\mathbf{D}(T) = (4 - \pi) \cdot n - \sqrt{\pi n} + \left(2 - \frac{\pi}{4}\right) - \sqrt{\frac{\pi}{n}} \cdot \frac{1}{8} - \frac{1}{32} \cdot \frac{\pi}{n} =$$

$$= 8{,}584 - 5{,}605 + 1{,}2146 - 0{,}0700 - 0{,}0098 = 4{,}1138 \ (\sqrt{\mathbf{D}(T)} = 2{,}028).$$

Видим, что значение T близко к $\mathbf{M}(T)$. Следовательно, можно ожидать принятия нулевой гипотезы. Вычисляем

$$M = \frac{2n - t}{2} = \frac{2 \cdot 10 - 5}{2} = 7{,}5; \quad x = \frac{4}{\sqrt{7{,}5}} = 1{,}4606;$$

$$\mathbf{P}(T > 5) = \sqrt{\frac{10}{7{,}5}} \cdot \left[1 - \frac{1}{96 \cdot 7{,}5} \cdot \left(1{,}4606^4 - 12 \cdot 1{,}4606^2 + 12\right)\right] \times$$

$$\times \left(1 + \frac{1}{80} + \frac{1}{128 \cdot 100}\right) \cdot e^{-\frac{1{,}4606^2}{4}} = 0{,}6938.$$

Отсюда $\mathbf{P}(T \leqslant 4) = 1 - 0{,}6938 \approx 0{,}3$, что существенно превышает уровень значимости $\alpha = 0{,}1$. Следовательно, нулевая гипотеза (совпадение эмпирических функций распределения вероятностей) не отклоняется.

3.1.2.8.3. Двухвыборочный критерий Андерсона

Поступим по аналогии с рассмотренным выше критерием (имеем выборку x объема n и выборку y объема m). Составим объединенную выборку и упорядочим ее по возрастанию. Обозначим через R_{x_i} и R_{y_i} — ранги элементов выборки x и y

в общем упорядоченном ряду (ранг — номер элемента, полученный им в упорядоченном ряду).

Статистика Андерсона имеет вид [237]

$$T = \frac{1}{mn(m+n)}\left[n\sum_{i=1}^{n}(R_{x_i} - i)^2 + m\sum_{j=1}^{m}(R_{y_j} - j)^2\right] - \frac{4mn-1}{6(m+n)}.$$

При $n, m \to \infty$ и $\frac{m}{n} = \text{const}$ статистика T распределена как статистика Смирнова–Крамера–фон Мизеса (см. раздел 3.1.2.2).

Напомним, что можно использовать критические точки из табл. 54.

Задача 103. *В условиях задачи* 102 *проверить нулевую гипотезу критерием Андерсона.*

Находим ранги R_{x_i} и R_{y_i} в общей выборке

$R_{x_1} = 3{,}5$; $R_{x_2} = 5$; $R_{x_3} = 7$; $R_{x_4} = 7$; $R_{x_5} = 9$; $R_{x_6} = 10$; $R_{x_7} = 12{,}5$;
$R_{x_8} = 14{,}5$; $R_{x_9} = 18$; $R_{x_{10}} = 19$; $R_{y_1} = 1$; $R_{y_2} = 2$; $R_{y_3} = 3{,}5$; $R_{y_4} = 7$;
$R_{y_5} = 11$; $R_{y_6} = 12{,}5$; $R_{y_7} = 14{,}5$; $R_{y_8} = 16$; $R_{y_9} = 17$; $R_{y_{10}} = 20$.

Если значения x-ов и y-ов совпадают, им присваиваются средние ранги. Например, значениям $x_3 = x_4 = y_4 = 14$ присваивается одинаковый средний ранг

$$R_{x_3} = R_{x_4} = R_{y_4} = \frac{6+7+8}{3} = 7.$$

По аналогии определяются все остальные ранги.

Вычисляем статистику критерия

$$T = \frac{1}{10\cdot 10\cdot(10+10)}\cdot\left[10\cdot\sum_{i=1}^{10}(R_{x_i} - i)^2 + 10\cdot\sum_{j=1}^{10}\left(R_{y_j} - 1\right)^2\right] - \frac{4\cdot 10\cdot 10 - 1}{6\cdot(10+10)} = 0{,}0675.$$

Из табл. 54 находим критическое значение (для $\alpha = 0{,}90$) $n\omega^2(\alpha) = 0{,}3473$. Так как $T = 0{,}0675 < 0{,}3473$, нулевая гипотеза не отклоняется.

3.2. Критерии нормальности распределения

Нормальный закон распределения вероятностей получил наибольшее распространение в практических задачах обработки экспериментальных данных. Большинство прикладных методов математической статистики исходит из предположения нормальности распределения вероятностей изучаемых случайных величин.

Широкое распространение этого распределения вызвало необходимость разработки специальных критериев согласия эмпирических распределений с нормальным.

3.2.1. Общие критерии согласия, модифицированные для проверки нормальности распределения

3.2.1.1. Модифицированный критерий χ^2

В [199] предложена и исследована форма критерия согласия χ^2, модифицированная применительно к проверке нормальности распределения, когда параметры распределения оцениваются по негруппированной выборке. После оценки параметров распределения совокупность выборочных данных разбивается на k равновероятных интервалов $\left(p_i = \dfrac{1}{k} = \text{const}\right)$ и статистика критерия подсчитывается по формуле

$$\chi^2 = \frac{k}{n}\sum_{i=1}^{k} m_i^2 - n,$$

где n — объем выборки; m_i — количество членов выборки, попавшее в i-й интервал.

Границы интервалов определяются как

$$\bar{x} + c_i s_i \ (i = 0, \ldots, k), \quad \text{где} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{и} \quad s = \left\{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{1}{2}}.$$

Значения коэффициентов c_i приведены в табл. 59. Следует помнить, что $c_0 = -\infty$ и $c_k = \infty$. Так как c_i симметричны относительно нуля, то недостающие значения c_i

Таблица 59

Значения коэффициентов c_i модифицированного χ^2-критерия нормальности для $k = 3 \div 15$ [199]

k	c_1	c_2	c_3	c_4	c_5	c_6	c_7
3	−0,4307						
4	−0,6745	0					
5	−0,8416	−0,2533					
6	−0,9674	−0,4307	0				
7	−1,0676	−0,5659	−0,1800				
8	−1,1503	−0,6745	−0,3186	0			
9	−1,2206	−0,7647	−0,4307	−0,1397			
10	−1,2816	−0,8416	−0,5244	−0,2533	0		
11	−1,3352	−0,9085	−0,6046	−0,3488	−0,1142		
12	−1,3830	−0,9674	−0,6745	−0,4307	−0,2194	0	
13	−1,4261	−1,0201	−0,7363	−0,5024	−0,2934	−0,0966	
14	−1,4652	−1,0676	−0,7916	−0,5660	−0,3661	−0,1800	0
15	−1,5011	−1,1108	−0,8416	−0,6229	−0,4307	−0,2533	−0,0837

можно найти из соотношений

$$c_{\frac{1}{2}(k-1)+i} = -c_{\frac{1}{2}(k-1)-i} \quad \left(i = 1, \ldots, \frac{k-1}{2}\right) \text{— для нечетных } k;$$

$$c_{\frac{1}{2}k+i} = -c_{\frac{1}{2}k-i} \quad \left(i = 1, \ldots, \frac{k-2}{2}\right) \text{— для четных } k.$$

Если $\chi^2 > d_k(\alpha)$, где $d_k(\alpha)$ — критическое значение статистики критерия на уровне значимости α, то гипотеза нормальности отклоняется. Критические значения $d_k(\alpha)$ приведены в табл. 60.

Таблица 60

Критические значения $d_k(\alpha)$ модифицированного χ^2-критерия нормальности

k	α			k	α		
	0,10	0,05	0,01		0,10	0,05	0,01
3	2,371	3,248	5,418	10	12,384	14,438	18,852
4	3,928	5,107	7,917	11	13,694	15,843	20,431
5	5,442	6,844	10,075	12	14,988	17,226	21,977
6	6,905	8,479	12,021	13	16,267	19,589	23,495
7	8,322	10,038	13,837	14	17,535	19,937	24,990
8	9,703	11,543	15,567	15	18,792	21,270	26,464
9	11,055	13,007	17,234				

Задача 104. *Для данных задачи 91 проверить модифицированным критерием χ^2 на уровне значимости $\alpha = 0,1$ гипотезу нормальности распределения при оценке его параметров по негруппированным данным.*

Имеем $\bar{x} = \frac{1}{100} \cdot \sum_{i=1}^{100} x_i = 100,77$; $s = \left\{\frac{1}{100} \cdot \sum_{i=1}^{100} (x_i - \bar{x})^2\right\}^{\frac{1}{2}} = 21,5827$.

Из табл. 59 находим коэффициенты разбиения (принимаем $k = 10$):

$$c_1 = -1,2816; \quad c_4 = -0,2533; \quad c_7 = 0,5244;$$
$$c_2 = -0,8416; \quad c_5 = 0; \quad c_8 = 0,8416;$$
$$c_3 = -0,5244; \quad c_6 = 0,2533; \quad c_9 = 1,2816.$$

Результаты расчетов сведем в таблицу:

i	Границы интервалов	m_i	m_i^2
1	$-\infty \div 72,79$	7	49
2	$72,79 \div 82,40$	12	144
3	$82,40 \div 89,32$	11	121
4	$89,32 \div 95,24$	11	121
5	$95,24 \div 100,77$	8	64
6	$100,77 \div 106,30$	13	169
7	$106,30 \div 112,22$	6	36
8	$112,22 \div 119,14$	12	144
9	$119,14 \div 128,40$	10	100
10	$128,40 \div \infty$	10	100
	$\sum$	100	1048

Статистика критерия равна

$$\chi^2 = \frac{k}{n} \cdot \sum_{i=1}^{k} m_i^2 - n = \frac{10}{100} \cdot 1048 - 100 = 4{,}8.$$

Из табл. 60 находим критическое значение статистики для $k = 10$ и $\alpha = 0{,}1$: $d_{10}(0{,}1) = 12{,}384$. Так как $\chi^2 = 4{,}8 < d_{10}(0{,}1) = 12{,}384$, гипотеза нормальности исходного распределения вероятностей не отклоняется.

3.2.1.2. Критерии типа Колмогорова–Смирнова

Применение критерия согласия $n\omega^2$ (см. раздел 1.2.2) для задачи проверки гипотезы нормальности распределения вероятностей случайных величин рассмотрено в [218]. Алгоритм вычисления статистики критерия в этом случае не меняется — меняются только критические значения статистики проверки гипотезы. Для различных ситуаций, когда параметры гипотетического распределения оцениваются непосредственно по самой выборке, критические значения статистики $n\omega^2$ приведены в табл. 61.

По аналогии в [209, 215] рассмотрено применение критерия Колмогорова–Смирнова (см. раздел 3.1.2.1) для проверки нормальности распределения в ситуации, когда оба его параметра оцениваются по выборке. Алгоритм проверки нулевой гипотезы H_0 и для этого случая сохраняется, меняются только критические значения — используется модифицированная статистика

$$D_n^{\text{н}} = D_n\left(\sqrt{n} - 0{,}01 + \frac{0{,}85}{\sqrt{n}}\right),$$

критические значения которой $D_n^{\text{н}}(\alpha)$ (α — уровень значимости) приведены в табл. 62.

Таблица 61

Критические значения статистики $n\omega^2$ для проверки нормальности распределения ($1-\alpha$ — уровень значимости) [218]

Исходные условия	α				
	0,90	0,95	0,99	0,995	0,999
Параметры (μ и σ) известны заранее	0,3473	0,4614	0,7435	0,8694	1,1679
Параметр σ известен, а параметр μ оценивается по выборке $\left(\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\right)$	0,1344	0,1653	0,2380	0,2698	0,3443
Параметр μ известен, а параметр σ оценивается по выборке $\left(s = \left\{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{1}{2}}\right)$	0,2370	0,4418	0,7245	0,8506	1,1490
Параметры μ и σ оцениваются по выборке	0,1035	0,1260	0,1788	0,2018	0,2559

Таблица 62

Критические значения статистики Колмогорова–Смирнова, модифицированной для проверки нормальности распределения [209]

α	0,15	0,10	0,05	0,03	0,01
$D_n^{\text{н}}(\alpha)$	0,775	0,819	0,895	0,955	1,035

Задача 105. *Для данных задачи 94 проверить гипотезу нормальности распределения случайных величин критерием типа Колмогорова–Смирнова с оценкой параметров распределения по выборке.*

Находим $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = 15{,}3$; $s = \left\{ \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}^{\frac{1}{2}} = 8{,}149.$

Критерий Колмогорова–Смирнова

Имеем $z_i = \frac{x_i - 15{,}3}{8{,}149}$. Результаты расчетов сведем в таблицу:

i	x_i	z_i	$\Phi(z_i)$	$\frac{i}{n}$	$\frac{i-1}{n}$	$\frac{i}{n} - \Phi(z_i)$	$\Phi(z_i) - \frac{i-1}{n}$
1	4	−1,386	0,0823	0,1	0,0	0,0179	0,0823
2	7	−1,018	0,1535	0,2	0,1	0,0465	0,0535
3	8	−0,896	0,1841	0,3	0,2	0,1159	−0,0159
4	9	−0,773	0,2207	0,4	0,3	0,1793	−0,0793
5	12	−0,405	0,3446	0,5	0,4	0,1554	−0,0554
6	18	0,331	0,6293	0,6	0,5	0,0293	−0,1293
7	19	0,454	0,6753	0,7	0,6	0,0247	−0,1293
8	21	0,699	0,7580	0,8	0,7	0,0420	−0,0420
9	25	1,190	0,8830	0,9	0,8	0,0170	−0,0170
10	30	1,804	0,9640	1,0	0,9	0,0360	−0,0360

Из таблицы следует, что

$$D_{10}^{+} = \max\left(\frac{i}{n} - \Phi(z_i)\right) = 0{,}1793; \quad D_{10}^{-} = \max\left(\Phi(z_i) - \frac{i-1}{n}\right) = 0{,}1293;$$

$$D_n = \max(D_{10}^{+}, D_{10}^{-}) = 0{,}1793.$$

Далее $D_n^{\text{н}} = 0{,}1793 \cdot \left(\sqrt{n} - 0{,}01 + \frac{0{,}85}{\sqrt{10}}\right) = 0{,}613$. Из табл. 62 имеем $D_n^{\text{н}}(\alpha) = 0{,}819$. Так как $D_n^{\text{н}} = 0{,}613 < D_n^{\text{н}}(0{,}1) = 0{,}819$, гипотеза нормальности распределения не отклоняется.

Критерий $n\omega^2$

Результаты работы сведены в таблицу:

i	x_i	z_i	$F(z_i)$	$F(z_i) - \frac{2i-1}{n}$	$\left\{F(z_i) - \frac{2i-1}{n}\right\}^2$
1	21	−1,386	0,0823	−0,0177	$3{,}13 \cdot 10^{-4}$
2	7	−1,018	0,1535	−0,1465	0,0214
3	8	−0,896	0,1841	−0,3159	0,0998
4	9	−0,773	0,2207	−0,4793	0,2297
5	12	−0,405	0,3446	−0,5554	0,3085
6	18	0,331	0,6293	−0,4707	0,2215
7	19	0,454	0,6753	−0,6247	0,3902
8	21	0,699	0,7580	−0,7420	0,5506
9	25	1,190	0,8830	−0,8170	0,6675
10	30	1,804	0,9640	−0,9360	0,8761

Находим $n\omega^2 = \frac{1}{12 \cdot 10} + 3{,}3656 = 3{,}374.$

Так как $n\omega^2 = 3{,}374 > n\omega^2(0{,}1) = 0{,}1035$ (см. табл. 61 при $1 - \alpha = 1 - 0{,}1 = 0{,}9$), нулевая гипотеза нормальности распределения отклоняется.

3.2.1.3. Критерий Фроцини

В [238, 239] Фроцини предложил простой, но достаточно мощный критерий нормальности с параметрами, оцениваемыми по выборке, основанный на статистике

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \Phi\left(z_i\right) - \frac{i-0{,}5}{n} \right|,$$

где $z_i = \dfrac{x_i - \bar{x}}{s}$; $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$; $s^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$; $\Phi(z_i)$ — функция распределения $N(0{,}1)$.

Критические значения статистики B_n приведены в табл. 63.

Таблица 63

Критические значения статистики Фроцини B_n для проверки нормальности распределения

(α — уровень значимости) [239]

n	α				
	0,80	0,85	0,90	0,95	0,99
5	0,2115	0,2245	0,2422	0,2666	0,3120
6	0,2132	0,2265	0,2434	0,2698	0,3148
7	0,2143	0,2276	0,2448	0,2702	0,3224
8	0,2175	0,2303	0,2477	0,2756	0,3286
9	0,2186	0,2314	0,2491	0,2753	0,3331
10	0,2179	0,2304	0,2485	0,2789	0,3332
11	0,2191	0,2332	0,2500	0,2774	0,3317
12	0,2212	0,2335	0,2508	0,2795	0,3356
13	0,2207	0,2336	0,2525	0,2784	0,3385
14	0,2204	0,2335	0,2518	0,2791	0,3367
15	0,2226	0,2354	0,2530	0,2820	0,3370
16	0,2226	0,2356	0,2519	0,2804	0,3376
17	0,2228	0,2362	0,2551	0,2812	0,3376
18	0,2233	0,2363	0,2536	0,2822	0,3374
19	0,2232	0,2367	0,2533	0,2830	0,3413
20	0,2257	0,2385	0,2556	0,2839	0,3363
∞	0,2250	0,2390	0,2560	0,2840	0,3410

Задача 106. *В условиях задачи 94 проверить гипотезу нормальности распределения случайных величин критерием Фроцини.*

Воспользуемся данными из таблицы расчета критерия Колмогорова–Смирнова и получим

$$B_n = \frac{1}{\sqrt{10}} \cdot (0{,}0323 + 0{,}0035 + 0{,}0659 + \\ + 0{,}1293 + 0{,}1054 + 0{,}0793 + 0{,}0253 + 0{,}008 + 0{,}033 + 0{,}014) = 0{,}1558.$$

Из табл. 63 находим критическое значение $B_n(\alpha = 0{,}1) = 0{,}2485$.

Так как $B_n = 0{,}1568 < B_n(0{,}1) = 0{,}2485$, гипотеза нормальности распределения случайных величин не отклоняется.

3.2.2. Специальные критерии нормальности

Учитывая чрезвычайно широкое распространение нормального распределения, предложено множество критериев проверки нормальности, использующих различные характеризации нормального распределения и направленные на защиту

Коэффициенты α_{n-i+1} $(\times 10^4)$

n	i											
	1	2	3	4	5	6	7	8	9	10	11	12
3	7071											
4	6872	1677										
5	6646	2413										
6	6431	2806	0875									
7	6233	3031	1401									
8	6052	3164	1743	0561								
9	5888	3244	1976	0947								
10	5739	3291	2141	1224	0399							
11	5601	3315	2260	1429	0695							
12	5475	3325	2347	1586	0922	0303						
13	5359	3325	2412	1707	1099	0539						
14	5251	3318	2460	1802	1240	0727	0240					
15	5150	3306	2495	1878	1353	0880	0433					
16	5056	3290	2521	1939	1447	1005	0593	0196				
17	4968	3273	2540	1988	1524	1109	0725	0359				
18	4886	3253	2553	2027	1587	1197	0837	0496	0173			
19	4808	3232	2561	2059	1641	1271	0932	0612	0303			
20	4734	3211	2565	2085	1686	1334	1013	0711	0422	0140		
21	4634	3185	2578	2119	1736	1399	1092	0804	0530	0263		
22	4590	3156	2571	2131	1764	143	1150	0878	0618	0368	0122	
23	4542	3126	2563	2139	1787	1480	1201	0941	0696	0459	0228	
24	4493	3098	2554	2124	1807	1512	1245	0997	0764	0539	0321	0107
25	4450	3069	2543	2148	1822	1539	1283	1046	0823	0610	0403	0200
26	4407	3043	2533	2151	1836	1563	1316	1089	0876	0672	0476	0284
27	4366	3018	2522	2152	1848	1584	1346	1128	0923	0728	0540	0358
28	4328	2992	2510	2151	1857	1601	1372	1162	0965	0778	0598	0424
29	4291	2968	2499	2150	1864	1616	1395	1192	1002	0822	0650	0483
30	4254	2944	2487	2148	1870	1630	1415	1219	1036	0862	0697	0537
31	4220	2921	2475	2145	1874	1641	1433	1243	1066	0899	0739	0585
32	4188	2898	2463	2141	1878	1651	1449	1265	1093	0931	0777	0629
33	4156	2876	2451	2137	1880	1660	1463	1284	1118	0961	0812	0669
34	4127	2854	2439	2132	1882	1667	1475	1301	1140	0988	0844	0706
35	4096	2834	2427	2127	1883	1673	1487	1317	1160	1013	0873	0739
36	4068	2813	2415	2121	1883	1678	1496	1331	1179	1036	0900	0770
37	4040	2794	2403	2116	1883	1683	1505	1344	1196	1056	0924	0798
38	4015	2774	2391	2110	1881	1686	1513	1356	1211	1075	0947	0824
39	3989	2755	2380	2104	1880	1689	1520	1366	1225	1092	0967	0848
40	3964	2737	2368	2098	1878	1691	1526	1376	1237	1108	0986	0870
41	3940	2719	2357	2091	1876	1693	1531	1384	1249	1123	1004	0891
42	3917	2701	2345	2085	1874	1694	1535	1392	1259	1136	1020	0909
43	3894	2684	2334	2078	1871	1695	1539	1398	1269	1149	1035	0927
44	3872	2667	2323	2072	1868	1695	1542	1405	1278	1160	1049	0943
45	3850	2651	2313	2065	1865	1695	1545	1410	1286	1170	1062	0959
46	3830	2635	2302	2058	1862	1695	1548	1415	1293	1180	1073	0972
47	3808	2620	2291	2052	1859	1695	1550	1420	1300	1189	1085	0986
48	3789	2604	2281	2045	1855	1693	1551	1423	1306	1197	1095	0998
49	3770	2589	2271	2038	1851	1692	1553	1427	1312	1205	1105	1010
50	3751	2574	2260	2032	1847	1691	1554	1430	1317	1212	1113	1020

Таблица 64

критерия Шапиро–Уилка [13, 240]

i												
13	14	15	16	17	18	19	20	21	22	23	24	25
0094												
0178												
0253	0084											
0320	0159											
0381	0227	0076										
0435	0289	0144										
0485	0344	0206	0068									
0530	0395	0262	0131									
0572	0441	0314	0187	0062								
0610	0484	0361	0239	0119								
0645	0523	0404	0287	0172	0057							
0677	0559	0444	0331	0220	0110							
0706	0592	0481	0372	0264	0158	0053						
0733	0622	0515	0409	0305	0203	0101						
0759	0651	0546	0444	0343	0244	0146	0049					
0782	0677	0575	0476	0379	0283	0188	0094					
0804	0701	0602	0506	0411	0318	0227	0136	0045				
0824	0724	0628	0534	0442	0352	0263	0175	0087				
0842	0745	0651	0560	0471	0383	0296	0211	0126	0042			
0860	0765	0673	0584	0497	0412	0328	0245	0163	0081			
0876	0783	0694	0607	0522	0439	0357	0277	0197	0118	0039		
0892	0801	0713	0628	0546	0465	0385	0307	0229	0153	0076		
0906	0817	0731	0648	0568	0489	0411	0335	0259	0185	0111	0037	
0919	0832	0748	0667	0588	0511	0436	0361	0288	0215	0143	0071	
0932	0846	0764	0685	0608	0532	0459	0386	0314	0244	0174	0104	0035

нулевой гипотезы нормальности распределения от всевозможных альтернатив. Мы попытались (насколько это возможно) сгруппировать эти критерии в относительно однородные группы.

3.2.2.1. Критерий Шапиро–Уилка

Критерий Шапиро–Уилка [240] основан на отношении оптимальной линейной несмещенной оценки дисперсии (см. раздел 2.1.2.1.6.6) к ее обычной оценке методом максимального правдоподобия (см. раздел 2.1.2.1.1). Статистика критерия имеет вид

$$W = \frac{1}{s^2}\left[\sum_{i=1}^{k} a_{n-i+1}(x_{n-i+1} - x_i)\right]^2, \quad \text{где} \quad s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2; \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Числитель является квадратом оценки среднеквадратического отклонения Ллойда [241].

Коэффициенты a_{n-i+1} приведены в табл. 64, заимствованной из [13, 240]. Критические значения статистики $W(\alpha)$ приведены в табл. 66, заимствованной из [13, 240].

Если $W < W(\alpha)$, то нулевая гипотеза нормальности распределения отклоняется на уровне значимости α. Приближенную вероятность получения эмпирического значения W при H_0 можно вычислить по формуле [242]

$$z = \gamma + \eta \ln\left(\frac{W - \varepsilon}{1 - W}\right),$$

где γ, η и ε — коэффициенты, приведенные в табл. 65.

Изучение мощности критерия Шапиро–Уилка [243] показало, что это — один из наиболее эффективных критериев проверки нормальности распределения случайных величин. Для больших n таблицы коэффициентов a_{n-i+1} становятся неудобными, поэтому была предложена модификация критерия Шапиро–Уилка — критерий Шапиро–Франчиа [244].

Его статистика имеет вид

$$W' = \frac{1}{s^2}\left[\sum_{i=1}^{k} c_{n-i+1}(x_{n-i+1} - x_i)\right]^2, \quad \text{где} \quad c_{n-i+1} = \frac{m_{n-i+1}}{\left(\sum_{i=1}^{n} m_{i,n}^2\right)^{\frac{1}{2}}}$$

и $m_{i,n}$ — математическое ожидание i-й порядковой статистики из стандартного нормального распределения. Аппроксимация $m_i = \Phi^{-1}\left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}}\right) = u_p$, где $p = \frac{i - \frac{3}{8}}{n + \frac{1}{4}}$, не искажает существенно критерий W' [245].

Используя аппроксимацию для квантили стандартного нормального распределения (см. раздел 1.1.1), можно записать

$$u_p = 4{,}91\left[p^{0,14} - (1-p)^{0,14}\right],$$

и для $p = \frac{i - \frac{3}{8}}{n + \frac{1}{4}}$ имеем $m_i = 4{,}91\left\{\left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}}\right)^{0,14} - \left(\frac{n - i + \frac{5}{8}}{n + \frac{1}{4}}\right)^{0,14}\right\}$.

В [36] приведена полезная аппроксимация, позволяющая применить критерий Шапиро–Уилка без помощи таблиц.

Таблица 65

Коэффициенты η, γ и ε [13]

n	γ	η	ε	n	γ	η	ε
3	−0,6250	0,3860	0,7500	27	−5,9050	1,9050	0,1980
4	−1,1070	0,7140	0,6297	28	−5,9880	1,9150	0,1943
5	−1,5300	0,9350	0,5521	29	−6,0740	1,9340	0,1907
6	−2,0100	1,1380	0,4963	30	−6,1600	1,9490	0,1872
7	−2,3560	1,2450	0,4533	31	−6,2480	1,9650	0,1840
8	−2,6960	1,3330	0,4186	32	−6,3240	1,9760	0,1811
9	−2,9680	1,4000	0,3900	33	−6,4020	1,9880	0,1781
10	−3,2620	1,4710	0,3660	34	−6,4800	2,0000	0,1755
11	−3,4850	1,5150	0,3451	35	−6,5590	2,0120	0,1727
12	−3,7310	1,5710	0,3270	36	−6,6400	2,0240	0,1702
13	−3,9360	1,6130	0,3111	37	−6,7210	2,0370	0,1677
14	−4,1550	1,6550	0,2969	38	−6,8030	2,0490	0,1656
15	−4,3730	1,6950	0,2842	39	−6,8870	2,0620	0,1633
16	−4,5670	1,7240	0,2727	40	−6,9610	2,0750	0,1612
17	−4,7130	1,7390	0,2622	41	−7,0350	2,0880	0,1591
18	−4,8850	1,7700	0,2528	42	−7,1110	2,1010	0,1572
19	−5,0180	1,7860	0,2440	43	−7,1880	2,1140	0,1552
20	−5,1840	1,7980	0,2375	44	−7,2120	2,1190	0,1548
21	−5,2910	1,8180	0,2264	45	−7,2660	2,1280	0,1534
22	−5,4130	1,8350	0,2207	46	−7,3450	2,1410	0,1526
23	−5,5080	1,8480	0,2157	47	−7,4140	2,1550	0,1499
24	−5,6050	1,8620	0,2106	48	−7,5550	2,1830	0,1466
25	−5,7040	1,8760	0,2063	49	−7,6150	2,1980	0,1451
26	−5,8030	1,8900	0,2020	50	−7,6770	2,2120	0,1436

Для $\alpha = 0{,}05$ предлагается статистика

$$W_1 = \left(1 - \frac{0{,}6695}{n^{0{,}6518}}\right)\frac{s^2}{B},$$

где

$$B = \left\{\sum_{j=1}^{m} a_j(x_{n-j} - x_j)\right\}^2; \quad m = \left[\frac{n}{2}\right]; \quad a_0 = \frac{0{,}899}{(n-2{,}4)^{0{,}4162}} - 0{,}02;$$

$$a_j = a_0\left[z + \frac{1483}{(3-z)^{10{,}845}} + \frac{71{,}610^{-10}}{(1{,}1-z)^{8{,}26}}\right]; \quad z = \frac{n-2j+1}{n-0{,}5}.$$

Если $W_1 < 1$, то нулевая гипотеза нормальности распределения случайных величин отклоняется. Модификация критерия Шапиро–Уилка для случая группированных данных (что существенно при наличии совпадающих наблюдений) рассмотрена в [246].

Задача 107. *Имеется выборка данных*

$$x:\ -1,\ 0,\ 1,\ 2,\ 3,\ 5,\ 6,\ 7,\ 10,\ 15.$$

Проверить гипотезу нормальности распределения случайной величины x критерием Шапиро–Уилка на уровне значимости $\alpha = 0{,}05$.

Таблица 66

Процентные точки критерия $W(\alpha)$ (α — уровень значимости) [13, 240]

n	α					n	α				
	0,01	0,02	0,05	0,10	0,50		0,01	0,02	0,05	0,10	0,50
3	0,737	0,756	0,767	0,789	0,959	27	0,894	0,906	0,923	0,935	0,965
4	0,687	0,707	0,748	0,792	0,935	28	0,896	0,908	0,924	0,936	0,966
5	0,686	0,715	0,762	0,806	0,927	29	0,898	0,910	0,926	0,937	0,966
6	0,713	0,743	0,788	0,826	0,927	30	0,900	0,912	0,927	0,939	0,967
7	0,730	0,760	0,803	0,838	0,928	31	0,902	0,914	0,929	0,940	0,967
8	0,749	0,778	0,818	0,851	0,932	32	0,904	0,915	0,930	0,941	0,968
9	0,764	0,791	0,829	0,859	0,935	33	0,906	0,917	0,931	0,942	0,968
10	0,781	0,806	0,842	0,869	0,938	34	0,908	0,919	0,933	0,943	0,969
11	0,792	0,817	0,850	0,876	0,940	35	0,910	0,920	0,934	0,944	0,969
12	0,805	0,828	0,859	0,883	0,943	36	0,912	0,922	0,935	0,945	0,970
13	0,814	0,837	0,866	0,889	0,945	37	0,914	0,924	0,936	0,946	0,970
14	0,825	0,846	0,974	0,895	0,947	38	0,916	0,925	0,938	0,947	0,970
15	0,835	0,855	0,881	0,901	0,950	39	0,917	0,927	0,939	0,948	0,971
16	0,844	0,863	0,887	0,906	0,952	40	0,919	0,928	0,940	0,949	0,971
17	0,851	0,869	0,892	0,910	0,954	41	0,920	0,929	0,941	0,950	0,972
18	0,858	0,874	0,897	0,914	0,956	42	0,922	0,930	0,942	0,951	0,972
19	0,863	0,879	0,901	0,917	0,957	43	0,923	0,932	0,943	0,951	0,972
20	0,868	0,884	0,905	0,920	0,959	44	0,924	0,933	0,944	0,952	0,973
21	0,873	0,888	0,908	0,923	0,960	45	0,926	0,934	0,945	0,953	0,973
22	0,878	0,892	0,911	0,926	0,961	46	0,927	0,935	0,945	0,953	0,973
23	0,881	0,895	0,914	0,928	0,962	47	0,928	0,936	0,946	0,954	0,974
24	0,884	0,889	0,916	0,930	0,963	48	0,929	0,937	0,947	0,954	0,974
25	0,888	0,901	0,918	0,931	0,964	49	0,929	0,937	0,947	0,955	0,974
26	0,891	0,904	0,920	0,933	0,965	50	0,930	0,938	0,947	0,955	0,974

Имеем $s^2 = \sum\limits_{i=1}^{n}(x_i - \bar{x})^2 = 219{,}6$. Используя коэффициенты a_{n-i+1} из табл. 64 для $n = 10$ $\left(k = \dfrac{n}{2} = 5\right)$, находим

$$\sum_{i=1}^{5} a_{n-i+1}\cdot(x_{n-i+1} - x_i) = 0{,}5739\cdot(15+1) + 0{,}3291\cdot(10-0) + 0{,}2141\cdot(7-1) + \\ + 0{,}1224\cdot(6-2) + 0{,}0399\cdot(5-3) = 14{,}3274;$$

$$\left[\sum_{i=1}^{5} a_{n-i+1}\cdot(x_{n-i+1} - x_i)\right]^2 = 205{,}27439; \quad W = \frac{205{,}27439}{219{,}6} = 0{,}935.$$

Из табл. 66 находим $W_{10}(0{,}1) = 0{,}869$.

Так как $W = 0{,}935 > W_{10}(0{,}1) = 0{,}869$, гипотеза нормальности распределения не отклоняется.

Вычислим точную вероятность получения значения $W = 0{,}935$ при условии справедливости нулевой гипотезы H_0.

Из табл. 65 для $n = 10$ находим $\gamma = -3{,}262$, $\eta = 1{,}471$ и $\varepsilon = 0{,}3660$.

Далее получаем $z = -3{,}262 + 1{,}471\cdot\ln\left(\dfrac{0{,}935 - 0{,}366}{1 - 0{,}935}\right) = -0{,}07076$. Соответствующая этой квантили стандартного нормального распределения вероятность $\Phi(-0{,}07067) \approx 0{,}47$ (см. табл. 1). Эта вероятность существенно превышает принятый уровень значимости $\alpha = 0{,}1$, что позволяет уверенно принять нулевую гипотезу нормальности.

Используем теперь критерий W', для чего предварительно вычислим (по табл. 1 или с помощью аппроксимаций)

$$i=1,\quad m_{10}=4{,}91\cdot\left(\left(\frac{10-\frac{3}{8}}{10+\frac{1}{4}}\right)^{0{,}14}-\left(\frac{10-10+\frac{5}{8}}{10+\frac{1}{4}}\right)^{0{,}14}\right)=1{,}5479;\quad i=2,\quad m_9=0{,}998;$$

$$i=3,\quad m_8=0{,}655;\quad i=4,\quad m_7=0{,}374;\quad i=5,\quad m_6=0{,}122;\quad m_7=-m_4;\quad m_8=-m_3;$$

$$m_9=-m_2;\quad m_{10}=-m_1;$$

$$c_1=\frac{m_{n-i+1}}{\left(\sum_{i=1}^{n}m_{i,n}^2\right)^{\frac{1}{2}}}=\frac{m_{10}}{2{,}8198}=\frac{1{,}5479}{2{,}8198}=0{,}5489;\quad c_2=0{,}3539;\quad c_3=0{,}2323;$$

$$c_4=0{,}132;\quad c_5=0{,}0433;\quad \sum_{i=1}^{5}c_{n-i+1}\cdot(x_{n-i+1}-x_i)=14{,}3322;\quad W'=\frac{205{,}36}{219{,}6}=0{,}935,$$

что очень близко к точному значению критерия.

В заключение используем приближенный критерий, не требующий применения таблиц. Имеем

$$a_0=\frac{0{,}899}{(10-2{,}4)^{0{,}4162}}-0{,}02=0{,}3665138;\quad z_j=\frac{10-2j+1}{9{,}5};$$

$$z_1=0{,}94737;\quad z_2=0{,}73684;\quad z_3=0{,}52616;\quad z_4=0{,}31579;\quad z_5=0{,}105263;$$

$$a_1=0{,}3665138\cdot\left[0{,}94737+\frac{1483}{(3-0{,}94737)^{10{,}845}}+\frac{71{,}6\cdot10^{-10}}{(1{,}1-0{,}94737)^{8{,}26}}\right]=0{,}584074;$$

$$a_2=0{,}3474026;\quad a_3=0{,}2222972;\quad a_4=0{,}1278953;\quad a_5=0{,}0439392;$$

$$B=[0{,}5847074\cdot(15+1)+0{,}347026\cdot(10-0)+0{,}2222972\cdot(7-1)+0{,}1278953\cdot(6-2)+ \\ +0{,}0439392\cdot(5-3)]^2=217{,}934;$$

$$W_1=\left(1-\frac{0{,}6695}{10^{0{,}6518}}\right)\cdot\frac{219{,}6}{217{,}934}=0{,}857.$$

Так как $W_1=0{,}857<1$, то нулевая гипотеза отклоняется на уровне значимости $\alpha=0{,}1$.

3.2.2.2. Энтропийный критерий нормальности (критерий Васичека)

Критерий основан на том, что энтропия нормального распределения превышает энтропию любого другого распределения с той же дисперсией.

Энтропия распределения вероятностей с плотностью $f(x)$ равна

$$H=-\int_{-\infty}^{\infty}f(x)\log f(x)\,dx,$$

а ее оценка по выборочным данным

$$H_{mn}=\frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{n}{2m}(x_{i+m}-x_{i-m})\right\},$$

где $x_i=x_1$ при $i<1$; $x_i=x_n$ при $i>n$ (x_i — i-я порядковая статистика), m — целое положительное число, меньшее, чем $n/2$.

Статистика критерия Васичека имеет вид [247]

$$K_{mn} = \frac{n}{2ms}\left\{\prod_{i=1}^{n}(x_{i+m} - x_{i-m})\right\}, \quad \text{где} \quad s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2; \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Если $K_{mn} < K_{mn}(\alpha)$, где $K_{mn}(\alpha)$ — критическое значение статистики, то нулевая гипотеза нормальности распределения отклоняется на уровне значимости α. Значения $K_{mn}(\alpha)$ для $\alpha = 0{,}05$ приведены в табл. 67. При $n, m \to \infty$, $m/n \to \infty$ и справедливости гипотезы H_0 $K_{mn} \to \sqrt{2\pi \cdot e} = 4{,}133$ и всегда $0 \leqslant K_{mn} \leqslant 4{,}133$.

Таблица 67

Значения $K_{mn}(\alpha)$ для $\alpha = 0{,}05$ [247]

n	m	K_{mn}	n	m	K_{mn}	n	m	K_{mn}	n	m	K_{mn}
3	1	0,99	9	3	2,13	18	2	2,62	30	5	3,05
4	1	1,05	10	1	1,76		3	2,69	35	2	3,00
5	1	1,19		2	2,15		4	2,67		3	3,13
	2	1,70		3	2,21	20	1	2,25		4	3,16
6	1	1,33	12	1	1,90		2	2,69		5	3,16
	2	1,77		2	2,31		3	2,77	40	3	3,19
7	1	1,46		3	2,36		4	2,76		4	3,24
	2	1,87	14	1	2,01	25	2	2,83		5	3,24
	3	1,87		2	2,43		3	2,93	45	3	3,25
8	1	1,57		3	2,49		4	2,93		4	3,29
	2	1,97	16	1	2,11		5	2,91		5	3,30
	3	2,05		2	2,54	30	2	2,93	50	3	3,29
9	1	1,67		3	2,60		3	3,04		4	3,34
	2	2,06	18	1	2,18		4	3,06		5	3,35

Критерий прост, не нуждается в таблице коэффициентов, как критерий Шапиро–Уилки (см. раздел 3.2.2.1), его асимптотическая эффективность удовлетворительна. Наиболее эффективен критерий Васичека при проверке нормальности распределения против альтернатив равномерности и экспоненциальности.

Исследования показали [248], что этот критерий чувствительнее к выбросам случайных величин, чем критерий Шапиро–Уилка.

Задача 108. *В условиях задачи* 107 *проверить нормальность распределения случайных величин критерием Васичека на уровне значимости*

$$\alpha = 0{,}05.$$

Для примера используем критерий при $m = 1$ и $m = 2$. Имеем

$$s = \left\{\frac{1}{n} \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{1}{2}} = 4{,}686,$$

и при $m = 1$ $(n = 10)$ получаем

$$K_{mn} = K_{1,10} = \frac{10}{2 \cdot 1 \cdot 4{,}686} \cdot \left\{\prod_{i=1}^{10}(x_{i+1} - x_{i-1})\right\}^{\frac{1}{10}} =$$

$$= 1{,}067 \cdot \left\{(x_2 - x_1) \cdot (x_3 - x_1) \cdot (x_4 - x_2) \ldots (x_{10} - x_8) \cdot (x_{10} - x_9)\right\}^{\frac{1}{10}} =$$

$$= 1{,}067 \cdot \left\{(0 + 1) \cdot (1 + 1) \cdot (2 - 0) \ldots (15 - 7) \cdot (15 - 10)\right\}^{\frac{1}{10}} = 2{,}913.$$

Из табл. 67 находим $K_{1,10}(0{,}05) = 1{,}76$.

Так как $K_{1,10} = 2{,}913 > K_{1,10}(0{,}05) = 1{,}76$, нулевая гипотеза нормальности распределения не отклоняется.

По аналогии имеем $K_{2,10} = 2{,}694$ и $K_{2,10}(0{,}05) = 2{,}15$.

Следовательно, и при $m = 2$ приходим к такому же результату.

3.2.2.3. Критерий Хегази–Грина

Хегази и Грин [249] рассмотрели серию критериев, основанных на статистиках

$$T_1 = \frac{1}{n}\sum_{i=1}^{n}\left|y_{(i)} - \eta_i\right|; \quad T_2 = \frac{1}{n}\sum_{i=1}^{n}\left\{y_{(i)} - \eta_i\right\}^2,$$

где $y_{(i)} = \dfrac{x_i - \bar{x}}{s}$; $s^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$; $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$; $\eta_i = \Phi^{-1}(m_i)$; $P_i = \dfrac{i}{n+1}$.

Напомним (см. раздел 1.1.1), что квантили u_{p_i} могут быть аппроксимированы по формуле

$$u_{p_i} = 4{,}91\left[p_i^{0,14} - (1-p_i)^{0,14}\right] = 4{,}91\left[\left(\frac{i}{n+1}\right)^{0,14} - \left(\frac{n+1-i}{n+1}\right)^{0,14}\right].$$

Критические значения величин $T_1(\alpha)$ и $T_2(\alpha)$ приведены в табл 68. Для $\alpha = 0{,}01$ и $\alpha = 0{,}05$ имеются весьма точные аппроксимации [249]:

$$T_1(0{,}01) = 0{,}7195 - 0{,}1751\ln n + 0{,}0108(\ln n)^2; \quad T_2(0{,}01) = 0{,}0178 + \frac{2{,}8736}{n} - \frac{8{,}2894}{n^2};$$

$$T_1(0{,}05) = 0{,}6027 - 0{,}1481\ln n + 0{,}0090(\ln n)^2; \quad T_2(0{,}05) = 0{,}0126 + \frac{1{,}9227}{n} + \frac{5{,}00677}{n^2}.$$

Этот критерий превосходит по мощности критерий Шапиро–Уилка (см. раздел 3.2.2.1) при альтернативах Коши и экспоненциального распределения, но уступает ему при равномерной альтернативе.

Таблица 68

Критические величины $T_1(\alpha)$ и $T_2(\alpha)$ (α — уровень значимости) [249]

n	α									
	0,01		0,025		0,05		0,10		0,50	
	$T_1(\alpha)$	$T_2(\alpha)$	$T_1(\alpha)$	$T_2(\alpha)$	$T_1(\alpha)$	$T_2(\alpha)$	$T_1(\alpha)$	$T_2(\alpha)$	$T_1(\alpha)$	$T_2(\alpha)$
5	0,4563	0,2616	0,4212	0,2285	0,3871	0,1974	0,3514	0,1683	0,3304	0,1509
10	0,3648	0,2180	0,3377	0,1834	0,3101	0,1529	0,2822	0,1278	0,2683	0,1129
20	0,2850	0,1462	0,2608	0,1181	0,2399	0,0988	0,2176	0,0819	0,2035	0,0718
40	0,2090	0,0873	0,1912	0,0712	0,1770	0,0599	0,1609	0,0490	0,1513	0,0429
80	0,1520	0,0476	0,1377	0,0392	0,1265	0,0322	0,1159	0,0274	0,1091	0,0239

Задача 109. *В условиях задачи* 107 *проверить гипотезу нормальности распределения критерием Хегази–Грина при* $\alpha = 0{,}05$.

Имеем

$$\bar{x} = 4{,}8; \quad s = 4{,}3996; \quad y_1 = \frac{x_1 - \bar{x}}{s} = \frac{-1 - 4{,}8}{4{,}9396} = -1{,}17418;$$

$$\eta_1 = U_{p_1} = 4{,}91 \cdot \left[\left(\frac{1}{10+1}\right)^{0,14} - \left(\frac{10+1-1}{10+1}\right)^{0,14}\right] = -1{,}33508.$$

Результаты расчетов сведем в таблицу:

i	x_i	y_i	p_i	η_i	$\lvert y_i - \eta_i \rvert$	$(y_i - \eta_i)^2$
1	−1	−1,17418	0,90900	−1,33508	0,16090	0,02589
2	0	−0,97170	0,18180	−0,90647	0,06523	$4{,}2559 \cdot 10^{-3}$
3	1	−0,76930	0,27270	−0,60250	0,16680	0,02780
4	2	−0,56680	0,36360	−0,34730	0,21950	0,04818
5	3	−0,36440	0,45450	−0,11370	0,25070	0,06285
6	5	0,04050	0,54540	−0,11370	0,07320	$5{,}3589 \cdot 10^{-3}$
7	6	0,24290	0,63630	0,34370	0,10080	0,01016
8	7	0,44540	0,72720	0,60250	0,15710	0,02468
9	10	1,05270	0,81820	0,90650	0,14620	0,02370
10	15	2,06490	0,90910	1,33510	0,72980	0,53260

Получаем $T_1 = \dfrac{2{,}657}{10} = 0{,}2657$ и $T_2 = \dfrac{0{,}763}{10} = 0{,}0763$.

Из табл. 68 для $\alpha = 0{,}05$ и $n = 10$ находим

$$T_1(0{,}05) = 0{,}3101; \quad T_2(0{,}05) = 0{,}1529.$$

Так как $T_1 = 0{,}2657 < T_1(0{,}05) = 0{,}3101$ и $T_2 = 0{,}0763 < T_2(0{,}05) = 0{,}1529$, гипотеза нормальности не отклоняется обоими критериями.

Проверим аппроксимацию для критических точек $T_1(0{,}05)$ и $T_2(0{,}05)$.

Имеем

$$T_1(0{,}05) = 0{,}6027 - 0{,}1481 \cdot \ln 10 + 0{,}0090 \cdot (\ln 10)^2 = 0{,}3094;$$

$$T_2(0{,}05) = 0{,}0126 + \frac{1{,}9227}{10} - \frac{5{,}0067}{100} = 0{,}1548.$$

Видим, что аппроксимации дают достаточно точные приближения к табличным значениям даже при относительно небольших объемах выборок n.

3.2.2.4. Критерий Али—Чёрго–Ревеса

Али, Чёрго и Ревес [250–252] предложили семейство статистик для проверки нормальности распределения случайных величин, основанных на взвешенных квадратах спейсингов (*спейсинг* — расстояние между соседними порядковыми статистиками).

Таблица 69

Критические значения $M_n(\alpha)$ критерия нормальности Али–Чёрго–Ревеса [252]

k	α			k	α		
	0,10	0,05	0,01		0,10	0,05	0,01
3	0,182	0,195	0,206	16	0,348	0,420	0,580
4	0,231	0,268	0,336	17	0,357	0,417	0,551
5	0,268	0,313	0,407	18	0,355	0,423	0,580
6	0,291	0,336	0,436	19	0,352	0,422	0,584
7	0,312	0,370	0,502	20	0,351	0,414	0,545
8	0,317	0,374	0,487	30	0,362	0,427	0,561
9	0,329	0,394	0,535	40	0,363	0,425	0,593
10	0,327	0,378	0,521	50	0,351	0,426	0,561
11	0,333	0,387	0,510	60	0,355	0,420	0,598
12	0,339	0,397	0,553	70	0,357	0,424	0,567
13	0,350	0,417	0,538	80	0,364	0,429	0,611
14	0,350	0,407	0,544	90	0,363	0,423	0,604
15	0,348	0,414	0,582	100	0,359	0,424	0,599

Если $x_1, \ldots, x_n$ — порядковые статистики наблюдаемого ряда случайных величин, то статистика критерия записывается в форме

$$M_n = \sum_{i=1}^{n} \left\{ \frac{x_n - \bar{x}}{s_n} - \Phi^{-1}\left(\frac{i}{n+1}\right) \right\}^2 \varphi\left[\Phi^{-1}\left(\frac{i}{n+1}\right)\right],$$

где $\Phi^{-1}(p) = u_p$ — p-квантиль стандартного нормального распределения; $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ — оценки соответственно среднего значения и дисперсии; $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$.

Критические значения $M_n(\alpha)$ для $\alpha = 0{,}01$; 0,05 и 0,10 приведены в табл. 69. Если $M_n < M_n(\alpha)$, то гипотеза нормальности распределения отклоняется. Мощность M_n-критерия для большинства возможных альтернатив не уступает мощности критериев Шапиро–Уилка (см. раздел 3.2.2.1) и Андерсона–Дарлинга (см. раздел 3.1.2.4).

Задача 110. *В условиях задачи* 107 *проверить гипотезу нормальности распределения случайных величин* M_n *критерием Али–Чёрго–Ревеса при* $\alpha = 0{,}05$.

Вычисления сведем в таблицу:

i	x_i	x_i^*	$\frac{i}{n+1}$	Φ_i^{-1}	φ_i	$x_i^* - \Phi_i^{-1}$	$(x_i^* - \Phi_i^{-1})\varphi_i$
1	−1	−1,158	0,091	−1,33	0,1647	0,0296	$4{,}875 \cdot 10^{-3}$
2	0	−0,955	0,182	−0,91	0,2637	0,00202	$5{,}34 \cdot 10^{-4}$
3	1	−0,751	0,273	−0,60	0,3332	0,0228	$7{,}597 \cdot 10^{-3}$
4	2	−0,548	0,364	−0,35	0,3752	0,0392	0,0147
5	3	−0,142	0,454	−0,11	0,3965	$1{,}02 \cdot 10^{-3}$	$4{,}06 \cdot 10^{-4}$
6	5	0,061	0,545	0,11	0,3965	$2{,}40 \cdot 10^{-3}$	$9{,}52 \cdot 10^{-4}$
7	6	0,264	0,636	0,35	0,3572	$7{,}396 \cdot 10^{-3}$	$2{,}774 \cdot 10^{-3}$
8	7	0,467	0,727	0,60	0,3332	0,0177	$5{,}894 \cdot 10^{-3}$
9	10	1,076	0,818	0,91	0,2637	0,0275	$7{,}266 \cdot 10^{-3}$
10	15	2,092	0,909	1,33	0,1647	0,5806	0,0956
						$\sum$	0,1406

В таблице для удобства записи приняты обозначения

$$x_i^* = \frac{x_i - \bar{x}}{s}; \quad \Phi^{-1}\left(\frac{i}{n+1}\right); \quad \varphi_i = \varphi\left[\Phi^{-1}\left(\frac{i}{n+1}\right)\right].$$

Из табл. 69 для $n = 10$ и $\alpha = 0{,}05$ имеем $M_{10}(0{,}05) = 0{,}378$.

Так как $M_n = 0{,}1406 < M_{10}(0{,}05) = 0{,}378$, гипотеза нормальности распределения случайных величин отклоняется.

3.2.2.5. Корреляционный критерий Филлибена

Филлибен [253] предложил критерий для проверки сложной гипотезы нормальности (когда параметры распределения не определены), статистикой которого является коэффициент корреляции r между порядковыми статистиками эмпирического ряда наблюдений x_i и медианами порядковых статистик стандартного нормального распределения (M_i)

$$r = \sum_{i=1}^{n}(x_i - \bar{x})(M_i - \bar{M}) \left\{ \sum_{i=1}^{n}(x_i - \bar{x})^2 (M_i - \bar{M})^2 \right\}^{\frac{1}{2}},$$

где $M_i = \Phi^{-1}(m_i)$ (Φ — функция Лапласа); m_i — медиана i-й порядковой статистики из равномерного распределения на интервале $[0, 1]$.

Так как $\bar{M} = 0$ и $M_i = -M_{n-i+1}$, то

$$r = \sum_{i=1}^{n} M_i x_i \left\{ \sum_{i=-1}^{n} M_i^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}^{-\frac{1}{2}}.$$

Более проста для вычислений статистики критерия формула

$$r = \sum_{i=\left[\frac{n}{2}\right]+1}^{n} M_i(x_i - x_{n-i+1}) \left\{ 2 \sum_{i=\left[\frac{n}{2}\right]+1}^{n} M_i^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}^{-\frac{1}{2}}.$$

Для вычисления m_i рекомендуется формула [253]

$$m_i = \begin{cases} i - m_n, & i = 1; \\ \dfrac{i - 0{,}375}{n + 0{,}365}, & i = 2, \ldots, n-1; \\ 0{,}5^{\frac{1}{n}}, & i = n. \end{cases}$$

В [254] показано, что использование при $i = 2, \ldots, n-1$ упрощенной формулы $m_i = \dfrac{i - 0{,}3}{n + 0{,}4}$ не изменяет существенно результаты. Критические значения критерия $r(\alpha)$ приведены в табл. 70.

Нулевая гипотеза принимается, если $r(\alpha) < r < r(1-\alpha)$. Критерий r имеет высокую мощность против симметричных альтернатив с длинными „хвостами“, по мощности сравним с критерием Шапиро–Уилка (см. раздел 3.2.2.1). Основные достоинства критерия — возможность проверки сложных гипотез, простота, легкость вычислений (нет необходимости иметь таблицу коэффициентов), достаточно высокая мощность. Критерий может быть распространен и на распределения, отличные от нормального (тогда $M_i = F^{-1}(m_i)$, где F — гипотетическое распределение), и в этом смысле является общим критерием согласия (см. раздел 3.1). Проверка гипотезы производится так же, как и для критерия Шапиро–Уилка.

Задача 111. *Проверить гипотезу нормальности распределения в условиях задачи* 107 *критерием Филлибена при* $\alpha = 0{,}05$.

Вычисляем

$$m_1 = \frac{i - 0{,}375}{n + 0{,}365} = \frac{1 - 0{,}375}{10 + 0{,}365} = 0{,}0603; \quad m_2 = 0{,}15678; \quad m_3 = 0{,}25326;$$

$$m_4 = 0{,}34973; \quad m_5 = 0{,}44621; \quad m_6 = 0{,}54269; \quad m_7 = 0{,}63917; \quad m_8 = 0{,}73565;$$
$$m_9 = 0{,}83213; \quad m_{10} = 0{,}92860.$$

Используя аппроксимацию

$$M_i = \Phi^{-1}(m_i) = U_{m_i} = 4{,}91 \cdot \left[m_i^{0{,}14} - (1 - m_i)^{0{,}14} \right],$$

имеем

$$M_1 = 4{,}91 \cdot (0{,}0603^{0{,}14} - 0{,}9397^{0{,}14}) = -1{,}55363; \quad M_2 = -1{,}0061;$$
$$M_3 = -0{,}66213; \quad M_4 = -0{,}38448; \quad M_5 = -0{,}13464; \quad M_6 = 0{,}10673;$$
$$M_7 = 0{,}35477; \quad M_8 = 0{,}62788; \quad M_9 = 0{,}96076; \quad M_{10} = 1{,}46622.$$

Таблица 70

Критические значения $r(\alpha)$ критерия Филлибена [253]

n	α					
	0,01	0,05	0,10	0,90	0,95	0,99
3	0,869	0,879	0,891	0,999	1,000	1,000
4	0,822	0,868	0,894	0,992	0,996	0,999
5	0,822	0,879	0,902	0,988	0,992	0,997
6	0,835	0,890	0,911	0,986	0,990	0,996
7	0,847	0,899	0,916	0,986	0,990	0,995
8	0,859	0,905	0,924	0,986	0,990	0,995
9	0,868	0,912	0,929	0,987	0,990	0,994
10	0,876	0,917	0,934	0,987	0,990	0,994
11	0,883	0,922	0,938	0,988	0,990	0,994
12	0,889	0,926	0,941	0,988	0,990	0,994
13	0,885	0,931	0,944	0,988	0,991	0,994
14	0,901	0,934	0,947	0,989	0,991	0,994
15	0,907	0,937	0,950	0,989	0,991	0,994
16	0,912	0,940	0,952	0,989	0,991	0,994
17	0,916	0,942	0,954	0,990	0,992	0,994
18	0,919	0,945	0,956	0,990	0,992	0,995
19	0,923	0,947	0,958	0,990	0,992	0,995
20	0,925	0,950	0,960	0,991	0,992	0,995
21	0,928	0,952	0,961	0,991	0,993	0,995
22	0,930	0,952	0,961	0,991	0,993	0,995
23	0,933	0,955	0,964	0,991	0,993	0,995
24	0,936	0,957	0,965	0,992	0,993	0,995
25	0,937	0,958	0,966	0,992	0,993	0,995
26	0,939	0,959	0,967	0,992	0,993	0,995
27	0,941	0,960	0,968	0,992	0,994	0,995
28	0,943	0,962	0,969	0,992	0,994	0,995
29	0,945	0,962	0,969	0,992	0,994	0,995
30	0,947	0,964	0,970	0,993	0,994	0,996
31	0,948	0,965	0,971	0,993	0,994	0,996

n	α					
	0,01	0,05	0,10	0,90	0,95	0,99
32	0,949	0,966	0,972	0,993	0,994	0,996
33	0,950	0,967	0,973	0,993	0,994	0,996
34	0,951	0,967	0,973	0,993	0,994	0,996
35	0,952	0,968	0,974	0,993	0,995	0,996
36	0,953	0,968	0,974	0,994	0,995	0,996
37	0,955	0,969	0,975	0,994	0,995	0,996
38	0,956	0,970	0,974	0,994	0,996	0,996
39	0,957	0,971	0,976	0,994	0,995	0,996
40	0,958	0,972	0,977	0,994	0,995	0,996
41	0,958	0,972	0,977	0,994	0,995	0,996
42	0,958	0,973	0,978	0,994	0,995	0,997
43	0,959	0,973	0,978	0,994	0,995	0,997
44	0,960	0,973	0,978	0,994	0,995	0,997
45	0,961	0,974	0,978	0,994	0,995	0,997
46	0,962	0,974	0,979	0,995	0,995	0,997
47	0,963	0,974	0,979	0,995	0,995	0,997
48	0,963	0,975	0,980	0,995	0,996	0,997
49	0,964	0,977	0,980	0,995	0,996	0,997
50	0,965	0,977	0,981	0,995	0,996	0,997
55	0,967	0,978	0,982	0,995	0,996	0,997
60	0,970	0,980	0,984	0,995	0,996	0,997
65	0,972	0,981	0,984	0,996	0,996	0,997
70	0,974	0,982	0,985	0,996	0,997	0,998
75	0,975	0,983	0,986	0,996	0,997	0,998
80	0,976	0,984	0,987	0,996	0,997	0,998
85	0,977	0,985	0,987	0,997	0,997	0,998
90	0,978	0,985	0,988	0,997	0,997	0,998
100	0,981	0,987	0,989	0,997	0,998	0,998

Далее вычисляем

$$\bar{x} = 4{,}8; \quad \sum_{i=1}^{n}(x_i - \bar{x})^2 = 219{,}6; \quad \sum_{i=1}^{10} M_i^2 = 7{,}63472; \quad \sum_{i=1}^{n} x_i \cdot M_i = 38{,}37695; \quad r = 0{,}93725.$$

Из табл. 70 находим для $\alpha = 0{,}05$ $r(0{,}05) = 0{,}917$ и $r(0{,}95) = 0{,}99$.

Так как $r(0{,}95) = 0{,}99 > r = 0{,}937 > r(0{,}05) = 0{,}917$, гипотеза нормальности не отклоняется.

3.2.2.6. Регрессионный критерий нормальности Ла Брека

В [255] Ла Брек рассмотрел серию критериев, основанных на оценке отклонения от линейности зависимости $M(x_i) = \mu + \sigma M(y_i)$, где x_i — порядковая статистика эмпирического распределения, $y_i = \dfrac{x_i - \mu}{\sigma}$ — порядковая статистика стандартного нормального распределения.

Критерий позволяет проверять сложные гипотезы нормальности (когда знание μ и σ не требуется).

Статистики критерия имеют вид

$$K_1 = \begin{cases} \dfrac{1}{s^2}\left\{\sum\limits_{i=1}^{[\frac{n}{2}]} a_i(x_i + x_{n-i+1})\right\}^2, & n = 2, 4, \ldots, 2k; \\ \dfrac{1}{s^2}\left\{\sum\limits_{i=1}^{[\frac{n}{2}]} b_i(x_i + x_{n-i+1}) + a_{\frac{n+1}{2}} x_{\frac{n+1}{2}}\right\}^2, & n = 1, 3, \ldots, 2k-1; \end{cases}$$

$$K_2 = \frac{1}{s^2}\left\{\sum_{i=1}^{[\frac{n}{2}]} b_i(x_i + x_{n-i+1})\right\}^2, \quad n = 1, 3, \ldots, 2k-1; \quad b_{\frac{n+1}{2}} = 0;$$

$$K_3 = \frac{K_1 + K_2}{2}, \quad \text{где} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Необходимо помнить, что $a_i = a_{n-i+1}$ и $b_i = b_{n-i+1}$.

Критерий со статистикой K_1 более чувствителен к асимметричным альтернативам (K_1 — это, по существу, квадратичное отклонение от гипотетической линейной зависимости). Критерий со статистикой K_2 более чувствителен к симметричным альтернативам (он построен на анализе кубических отклонений от линейной зависимости). Комбинированный критерий K_3 является комбинацией K_1 и K_2 и применим при промежуточных альтернативах.

Коэффициенты a_i и b_i и критические значения статистик $K_1(\alpha)$, $K_2(\alpha)$ и $K_3(\alpha)$ приведены соответственно в табл. 71 и 72.

При $n > 12$ применимы аппроксимации для критических значений статистик:

$$K_1(0{,}90) = 2{,}682 + \frac{0{,}1450}{\sqrt{n}}; \quad K_1(0{,}95) = 3{,}822 - \frac{0{,}1809}{\sqrt{n}}; \quad K_1(0{,}99) = 6{,}900 - \frac{2{,}948}{\sqrt{n}};$$

$$K_2(0{,}90) = 2{,}667 - \frac{0{,}1937}{\sqrt{n}}; \quad K_2(0{,}95) = 3{,}815 - \frac{0{,}4944}{\sqrt{n}}; \quad K_2(0{,}99) = 6{,}893 - \frac{1{,}087}{\sqrt{n}};$$

$$K_3(0{,}90) = 2{,}331 - \frac{0{,}566}{\sqrt{n}}; \quad K_3(0{,}95) = 2{,}992 - \frac{0{,}3591}{\sqrt{n}}; \quad K_3(0{,}99) = 4{,}618 - \frac{1{,}067}{\sqrt{n}}.$$

Мощность критериев K_1, K_2, K_3 не уступает мощности критерия Шапиро–Уилка (см. раздел 3.2.2.1), превосходя его в среднем. Критерии являются односторонними, т. е. нулевая гипотеза нормальности отклоняется, если $K_i > K_i(\alpha)$. Более мощен критерий K_3.

Таблица 71

Коэффициенты a_i и b_i для вычисления статистик K_1, K_2, K_3 [255]

n	i	a_i	b_i	n	i	a_i	b_i	n	i	a_i	b_i
4	1	1,1387	−0,6261	14	1	1,1966	−1,2814	19	5	−0,1644	0,5153
	2	−1,1387	2,1699		2	0,2100	0,6499		6	−0,2329	0,4354
5	1	1,2002	−0,8403		3	−0,6730	0,7313		7	−0,0281	0,3377
	2	−0,7570	1,9742		4	−0,2261	0,6438		8	−0,3119	0,2298
	3	−0,9689	0		5	−0,3237	0,4888		9	−0,3300	0,1162
6	1	1,2266	−0,9701		6	−0,3813	0,3027		10	−0,3358	0
	2	−0,4545	1,7195		7	−0,4082	0,1023	20	1	1,1436	−1,3246
	3	−0,7721	0,6243	15	1	1,1875	−1,2927		2	0,3367	0,3380
7	1	1,2367	−1,0574		2	0,2393	0,5834		3	0,1002	0,5156
	2	−0,2768	1,4962		3	−0,0299	0,6906		4	−0,0433	0,5437
	3	−0,6095	0,8410		4	−0,1865	0,6327		5	−0,1406	0,5085
	4	−0,7009	0		5	−0,2854	0,5079		6	−0,2093	0,4413
8	1	1,2385	−1,1196		6	−0,3473	0,3512	20	7	−0,2583	0,3560
	2	−0,1483	1,3092		7	−0,3815	0,1789		8	−0,2920	0,2603
	3	−0,4796	0,9090		8	−0,3925	0		9	−0,3134	0,1583
	4	−0,6106	0,3174	16	1	1,1785	−1,3020		10	−0,3237	0,0531
9	1	1,2298	−1,1196		2	0,2645	0,5240	21	1	1,1363	−1,3283
	2	−0,0512	1,1527		3	0,0027	0,6517		2	0,3447	0,3041
	3	−0,3753	0,9145		4	−0,1513	0,6178		3	0,1223	0,4834
	4	−0,5245	0,4864		5	−0,2507	0,5177		4	−0,0216	0,5242
	5	−0,5691	0		6	−0,3154	0,3854		5	−0,1184	0,5002
10	1	1,2298	−1,2010		7	−0,3548	0,2366		6	−0,1876	0,4440
	2	0,0245	1,0205		8	−0,3736	0,0797		7	−0,2376	0,3695
	3	−0,2902	0,8919	17	1	1,1685	−1,3095		8	−0,2730	0,2844
	4	−0,4478	0,5771		2	0,2863	0,4705		9	−0,2696	0,1926
	5	−0,5164	0,1979		3	0,0315	0,6147		10	−0,3107	0,0972
11	1	1,2225	−1,2284		4	−0,1199	0,6006		11	−0,3152	0
	2	0,0852	0,9075		5	−0,2192	0,5210	22	1	1,1283	−1,3309
	3	−0,2198	0,8564		6	−0,2858	0,4092		2	0,3561	0,2705
	4	−0,3807	0,6236		7	−0,3289	0,2802		3	0,1390	0,4556
	5	−0,4629	0,3239		8	−0,3532	0,1421		4	−0,0023	0,5051
	6	−0,4886	0		9	−0,3611	0		5	−0,0984	0,4906
12	1	1,2143	−1,2501	18	1	1,1607	−1,3156		6	−0,1673	0,4439
	2	0,1346	0,8100		2	0,3053	0,4222		7	−0,2179	0,3791
	3	−0,1606	0,8157		3	0,0571	0,5797		8	−0,2548	0,3031
	4	−0,3222	0,6441		4	−0,0917	0,5822		9	−0,2805	0,2207
	5	−0,4125	0,4048		5	−0,1905	0,5198		10	−0,2971	0,1336
	6	−0,4536	0,1375		6	−0,2583	0,4252		11	−0,3051	0,0451
13	1	1,2956	−1,2674		7	−0,3040	0,3130	23	1	1,2050	−1,3329
	2	0,1756	0,7249		8	−0,3324	0,1911		2	0,3663	0,2395
	3	−0,1104	0,7733		9	−0,3461	0,0642		3	0,1543	0,4292
	4	−0,2710	0,6489	19	1	1,1521	−1,3206		4	0,0151	0,4864
	5	−0,3660	0,4566		2	0,3220	0,3782		5	−0,0798	0,4801
	6	−0,4171	0,2342		3	0,0798	0,5467		6	−0,1485	0,4419
	7	−0,4333	0		4	−0,0663	0,5630		7	−0,1995	0,3854

Продолжение таблицы 71

n	i	a_i	b_i	n	i	a_i	b_i	n	i	a_i	b_i
23	8	−0,2373	0,3178	27	9	−0,2071	0,2984	32	2	0,4232	0,0390
	9	−0,2648	0,2432		10	−0,2307	0,2434		3	0,2456	0,2468
	10	−0,2833	0,1644		11	−0,2483	0,1849		4	0,1237	0,3420
	11	−0,2942	0,0827		12	−0,2603	0,1246		5	0,0384	0,3789
	12	−0,2977	0		13	−0,2674	0,0625		6	−0,0259	0,3855
24	1	1,1129	−1,3345		14	−0,2697	0		7	−0,0759	0,2744
	2	0,3754	0,2107	28	1	1,0845	−1,3369		8	−0,1159	0,3519
	3	0,1681	0,4044		2	0,4038	0,1140		9	−0,1483	0,3217
	4	0,0312	0,4682		3	0,2130	0,3176		10	−0,1743	0,2863
	5	−0,0626	0,4692		4	0,0841	0,4007		11	−0,1953	0,2471
	6	−0,1310	0,4384		5	−0,0054	0,4235		12	−0,2220	0,2051
	7	−0,1822	0,3893		6	−0,0717	0,4152		13	−0,2247	0,1614
	8	−0,2208	0,3291		7	−0,1229	0,3888		14	−0,2341	0,1161
	9	−0,2494	0,2618		8	−0,1628	0,3512		15	−0,2402	0,0701
	10	−0,2698	0,1896		9	−0,1941	0,3060		16	−0,2432	0,0231
	11	−0,2828	0,1148		10	−0,2186	0,2556	34	1	1,0469	−1,3335
	12	−0,2891	0,0386		11	−0,2371	0,2017		2	0,4305	0,0075
25	1	1,1056	−1,3356		12	−0,2503	0,1459		3	0,2587	0,2162
	2	0,3836	0,1840		13	−0,2591	0,0878		4	0,1399	0,3156
	3	0,1808	0,3809		14	−0,2633	0,0295		5	0,0565	0,3577
	4	0,0460	0,4505	29	1	1,0779	−1,3367		6	−0,0066	0,3701
	5	−0,0466	0,4579		2	0,4093	0,0935		7	−0,5600	0,3647
	6	−0,1147	0,4337		3	0,2221	0,2985		8	−0,0957	0,3481
	7	−0,1661	0,3913		4	0,0949	0,3853		9	−0,1284	0,3236
	8	−0,2050	0,3377		5	0,0068	0,4120		10	−0,1547	0,2940
	9	−0,2347	0,2766		6	−0,0593	0,4081		11	−0,1765	0,2604
	10	−0,2563	0,2111		7	−0,1102	0,3861		12	−0,1941	0,2241
	11	−0,2713	0,1420		8	−0,1501	0,3528		13	−0,2082	0,1855
	12	−0,2798	0,0715		9	−0,1820	0,3117		14	−0,2189	0,1459
	13	−0,2875	0		10	−0,2067	0,2659		15	−0,2269	0,1048
26	1	1,0984	−1,3363		11	−0,2263	0,2159		16	−0,2321	0,0632
	2	0,3910	0,1591		12	−0,2406	0,1639		17	−0,2347	0,0210
	3	0,1926	0,3586		13	−0,2505	0,1103	36	1	1,0356	−1,3313
	4	0,0596	0,4334		14	−0,2564	0,0553		2	0,4367	−0,0209
	5	−0,0318	0,4464		15	−0,2583	0		3	0,2702	0,1881
	6	−0,0995	0,4282	30	1	1,0714	−1,3364		4	0,1543	0,2910
	7	−0,1506	0,3918		2	0,4144	0,0743		5	0,0724	0,3376
	8	−0,1904	0,3438		3	0,2305	0,2805		6	0,0108	0,3545
	9	−0,2205	0,2888		4	0,1052	0,3703		7	−0,0382	0,3541
	10	−0,2434	0,2286		5	0,0180	0,4008		8	−0,0775	0,3424
	11	−0,2597	0,1653		6	−0,0474	0,4008		9	−0,1099	0,3230
	12	−0,2702	0,1000		7	−0,0980	0,3827		10	−0,1368	0,2982
	13	−0,2754	0,0335		8	−0,1382	0,3533		11	−0,1589	0,2996
27	1	1,0914	−1,3367		9	−0,1701	0,3163		12	−0,1772	0,2379
	2	0,3977	0,1358		10	−0,1955	0,2741		13	−0,1921	0,2043
	3	0,2031	0,3376		11	−0,2156	0,2281		14	−0,2042	0,1689
	4	0,0723	0,4168		12	−0,2308	0,1799		15	−0,2134	0,1326
	5	−0,0181	0,4349		13	−0,2419	0,1294		16	−0,2202	0,0953
	6	−0,0851	0,4220		14	−0,2491	0,0783		17	−0,2247	0,0573
	7	−0,1365	0,3908		15	−0,2527	0,0261		18	−0,2269	0,0193
	8	−0,1762	0,3483	32	1	1,0589	−1,3352	38	1	1,0248	−1,3287

Окончание таблицы 71

n	i	a_i	b_i	n	i	a_i	b_i	n	i	a_i	b_i
38	2	0,4420	−0,0465	42	14	−0,1632	0,2121	46	22	−0,1959	0,0381
	3	0,2803	0,1624		15	−0,1746	0,1865		23	−0,1971	0,0127
	4	0,1669	0,2681		16	−0,1847	0,1588	48	1	0,9770	−1,3129
	5	0,0870	0,3182		17	−0,1923	0,1315		2	0,4853	−0,1451
	6	0,0264	0,3393		18	−0,1986	0,1025		3	0,3160	0,0599
	7	−0,0219	0,3431		19	−0,2031	0,0736		4	0,2137	0,1735
	8	−0,0609	0,3356		20	−0,2061	0,0442		5	0,1410	0,2352
	9	−0,0934	0,3204		21	−0,2076	0,0150		6	0,0854	0,2695
	10	−0,1199	0,3000	44	1	0,9949	−1,3197		7	0,0403	0,2873
	11	−0,1425	0,2755		2	0,4534	−0,1107		8	0,0035	0,2942
	12	−0,1615	0,2479		3	0,3040	0,0965		9	−0,0275	0,2935
	13	−0,1769	0,2185		4	0,1978	0,2078		10	−0,0537	0,2875
	14	−0,1897	0,1874		5	0,1224	0,2660		11	−0,0767	0,2774
	15	−0,2003	0,1543		6	0,0648	0,2960		12	−0,0962	0,2643
	16	−0,2081	0,1213		7	0,0186	0,3093		13	−0,1129	0,2491
	17	−0,2142	0,0868		8	−0,0189	0,3115		14	−0,1281	0,2150
	18	−0,2180	0,0524		9	−0,0509	0,3061		15	−0,1408	0,2128
	19	−0,2199	0,1178		10	−0,0774	0,2953		16	−0,1520	0,1926
40	1	1,0144	−1,3259		11	−0,1001	0,2805		17	−0,1611	0,1724
	2	0,4465	−0,0699		12	−0,1195	0,2629		18	−0,1695	0,1502
	3	0,2890	0,1388		13	−0,1367	0,2424		19	−0,1762	0,1280
	4	0,1784	0,2466		14	−0,1506	0,2208		20	−0,1817	0,1053
	5	0,1001	0,2999		15	−0,1631	0,1971		21	−0,1859	0,0825
	6	0,0402	0,3245		16	0,1731	0,1730		22	−0,1891	0,0594
	7	−0,0071	0,3318		17	−0,1817	0,1476		23	−0,1914	0,0352
	8	−0,0457	0,3280		18	−0,1887	0,1215		24	−0,1924	0,0115
	9	0,0780	0,3166		19	−0,1941	0,0951	50	1	0,9686	−1,3094
	10	−0,1046	0,2998		20	−0,1982	0,0681		2	0,4602	−0,1605
	11	−0,1276	0,2788		21	−0,2008	0,0409		3	0,3209	0,0435
	12	−0,1463	0,2553		22	−0,2021	0,0141		4	0,2207	0,1577
	13	−0,1629	0,2288	46	1	0,9858	−1,3163		5	0,1490	0,2211
	14	−0,1760	0,2015		2	0,4560	−0,1286		6	0,0943	0,2570
	15	−0,1873	0,1721		3	0,3103	0,0775		7	0,0500	0,2767
	16	−0,1962	0,1422		4	0,2062	0,1901		8	0,0138	0,2855
	17	−0,2033	0,1110		5	0,1321	0,2502		9	0,0043	0,2869
	18	−0,2084	0,0800		6	0,0754	0,2825		10	−0,0431	0,2828
	19	−0,2118	0,0480		7	0,0301	0,2981		11	−0,0660	0,2747
	20	−0,2135	0,0164		8	−0,0074	0,3029		12	−0,0853	0,2637
42	1	1,0045	−1,3229		9	−0,0387	0,3000		13	−0,1027	0,2501
	2	0,4501	−0,0911		10	−0,0651	0,2918		14	−0,1175	0,2347
	3	0,2971	0,1167		11	−0,0881	0,2794		15	−0,1303	0,2182
	4	0,1885	0,2266		12	−0,1074	0,2642		16	−0,1419	0,1999
	5	0,1118	0,2825		13	−0,1245	0,2464		17	−0,1517	0,1809
	6	0,0530	0,3100		14	−0,1390	0,2270		18	−0,1601	0,1613
	7	0,0064	0,3205		15	−0,1516	0,2061		19	−0,1673	0,1409
	8	−0,0319	0,3199		16	−0,1624	0,1838		20	−0,1734	0,1200
	9	−0,0638	0,3117		17	−0,1713	0,1610		21	−0,1784	0,0985
	10	−0,0905	0,2981		18	−0,1789	0,1373		22	−0,1823	0,0767
	11	−0,1134	0,2804		19	−0,1852	0,1128		23	−0,1851	0,0555
	12	−0,1326	0,2599		20	−0,1899	0,0885		24	−0,1870	0,0331
	13	−0,1492	0,2369		21	−0,1935	0,0633		25	−0,1880	0,0112

Таблица 72

Критические точки критериев $K_1(\alpha)$, $K_2(\alpha)$ и $K_3(\alpha)$ [255]

n	$\alpha = 0{,}90$			$\alpha = 0{,}95$			$\alpha = 0{,}99$		
	K_1	K_2	K_3	K_1	K_2	K_3	K_1	K_2	K_3
4	2,730	3,419	4,379	2,274	3,373	5,215	2,029	2,405	2,939
5	2,739	3,562	5,008	2,490	3,425	5,247	2,080	2,652	3,731
6	2,739	3,638	5,367	2,581	3,475	5,430	2,109	2,769	4,165
7	2,736	3,684	5,594	2,621	3,518	5,635	2,128	2,829	4,424
8	2,733	3,713	5,748	2,638	3,554	5,825	2,131	2,865	4,995
9	2,730	3,732	5,859	2,664	3,583	5,990	2,142	2,872	4,974
10	2,726	3,745	5,941	2,644	3,608	6,130	2,152	2,879	4,955
11	2,723	3,755	6,005	2,641	3,628	6,249	2,160	2,884	4,940
12	2,724	3,770	6,049	2,612	3,673	6,579	2,167	2,889	4,296

Задача 112. *Проверить гипотезу нормальности распределения случайных величин в условиях задачи* 107 *регрессионным критерием Ла Брека на уровне достоверности* $\alpha = 0{,}95$.

Используя коэффициенты a_i и b_i из табл. 71, имеем

$$\sum_{i=1}^{5} a_i \cdot (x_i + x_{n-i+1}) = 1{,}2298 \cdot (-1+15) + 0{,}0245 \cdot (0+10) - 0{,}2902 \cdot (1+7) - {}$$
$$- 0{,}4478 \cdot (2+6) - 0{,}5164 \cdot (3+5) = 7{,}427; \quad \left\{\sum_{i=1}^{5} a_i \cdot (x_i + x_{n-i+1})\right\}^2 = 55{,}1603;$$

$$\sum_{i=1}^{5} b_i \cdot (x_i + x_{n-i+1}) = -1{,}201 \cdot (-1+15) + 1{,}0205 \cdot (0+10) + 0{,}8919 \cdot (1+7) + {}$$
$$+ 0{,}5771 \cdot (2+6) + 0{,}1979 \cdot (3+5) = 6{,}7262; \quad \left\{\sum_{i=1}^{5} b_i \cdot (x_i + x_{n-i+1})\right\}^2 = 45{,}241;$$

$$K_1 = \frac{55{,}1603}{24{,}4} = 2{,}260; \quad K_2 = \frac{45{,}241}{24{,}4} = 1{,}854; \quad K_3 = \frac{2{,}26 + 1{,}854}{2} = 2{,}057.$$

Из табл. 72 для $\alpha = 0{,}95$ и $n = 10$ находим критические значения

$$K_1(0{,}95) = 3{,}745; \quad K_2(0{,}95) = 3{,}608 \quad \text{и} \quad K_3(0{,}95) = 2{,}879.$$

Так как $K_1 = 2{,}26 < K_1(0{,}95) = 3{,}745$; $K_2 = 1{,}854 < K_2(0{,}95) = 3{,}608$ и $K_3 = 2{,}057 < K_3(0{,}95) = 2{,}879$, гипотеза нормальности распределения случайных величин не отклоняется.

Если бы мы воспользовались аппроксимациями, то получили бы

$$K_1(0{,}95) = 3{,}822 - \frac{0{,}1809}{\sqrt{10}} = 3{,}764; \quad K_2(0{,}95) = 3{,}815 - \frac{0{,}4944}{\sqrt{10}} = 3{,}659;$$
$$K_3 = 2{,}922 - \frac{0{,}3591}{\sqrt{10}} = 2{,}878.$$

Видим, что даже при $n = 10$ аппроксимации достаточно удовлетворительны.

3.2.2.7. Критерий нормальности Локка–Спурье

Весьма эффективный критерий нормальности против асимметричных альтернатив предложен в [256]. Его статистики имеют вид

$$T_{1n} = \frac{U_{1n}}{s}; \quad T_{2n} = \frac{U_{2n}}{s},$$

где

$$U_{1n} = \frac{1}{C_n^3} \sum_{1 \leqslant i < j < k \leqslant n} (x_k - 2x_j + x_i) = \frac{1}{C_n^3} \sum_{i=1}^{n} \omega_i x_i;$$

$$U_{2n} = \frac{1}{C_n^3} \sum_{1 \leqslant i < j < k \leqslant n} \left\{ (x_k - x_j)^2 - (x_j - x_i)^2 \right\} = \frac{1}{C_n^3} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} V_{ij}(x_j - x_i)^2,$$

где

$$\omega_i = C_{i-1}^2 - 2(n-1)(i-1) + C_{n-1}^2; \quad V_{ij} = i + j - n - 1;$$

x_i — i-я порядковая статистика.

Математические ожидания и дисперсии статистик равны:

$$\mathbf{M}(T_{1n}) = 0; \quad \mathbf{D}(T_{1n}) = \frac{1}{C_n^3} \left\{ 1{,}03804 + 0{,}69714(n-3) + 0{,}0890805(n-3)(n-4) \right\};$$

$$\mathbf{M}(T_{2n}) = 0; \quad \mathbf{D}(T_{2n}) = \frac{1}{C_n^3} \left\{ 7{,}03804 + 5{,}32251(n-3) + 0{,}74412(n-3)(n-4) \right\}.$$

При $n \geqslant 5$ случайная величина $T' = \dfrac{T_{1n} - \mathbf{M}(T_{1n})}{\sqrt{\mathbf{D}(T_{1n})}}$ распределена как нормальная случайная величина и, следовательно, ее критические значения равны $T_{1n} = \mathbf{M}(T_{1n}) + \sqrt{\mathbf{D}(T_{1n})}u_\alpha$, где u_α — α-квантиль стандартного нормального распределения.

По аналогии, нормальное приближение для T_{2n} применимо при $n \geqslant 10$ и имеет вид $T_{2n}(\alpha) = \mathbf{M}(T_{2n}) + \sqrt{\mathbf{D}(T_{2n})}u_\alpha$.

Критерий T_{2n} более эффективен для несимметричных альтернатив с „мягкими“ хвостами. Развитие этих критериев для альтернатив с обоими „хвостами“, отличающимися от нормального распределения, представлено в [257].

Гипотеза нормальности распределения вероятностей случайной величины отклоняется, если $T_{1n}(T_{2n}) > T_{1n}(\alpha)(T_{2n}(\alpha))$, где $(1-\alpha)$ — уровень значимости.

Задача 113. *Проверить гипотезу нормальности распределения случайных величин в условиях задачи* 107 *критерием Локка–Спурье на уровне значимости* $\alpha = 0{,}05$.

Имеем $s^2 = 24{,}4$ и $s = 4{,}9396$. Используем статистику T_{1n}.

Предварительно необходимо найти совокупность значений $(x_k - 2x_j + x_i)$ для всех возможных сочетаний i, j, k, отвечающих соотношению $1 \leqslant i < j < k \leqslant n$ (всего их $Cn^3 = C_{10}^3 = 120$). Порядок вычислений иллюстрируется в таблице (в ней принято обозначение $x_{kji} = x_k - 2x_j + x_i$):

i	j	k	x_k	x_j	x_i	x_{kji}	i	j	k	x_k	x_j	x_i	x_{kji}	i	j	k	x_k	x_j	x_i	x_{kji}
1	2	3	1	0	−1	0	1	3	8	7	1	−1	4	1	5	9	10	3	−1	3
1	2	4	2	0	−1	1	1	3	9	10	1	−1	7	1	5	10	15	3	−1	8
1	2	5	3	0	−1	2	1	3	10	15	1	−1	12	1	6	7	6	5	−1	−5
1	2	6	5	0	−1	4	1	4	5	3	2	−1	−2	1	6	8	7	5	−1	−4
1	2	7	6	0	−1	5	1	4	6	5	2	−1	0	1	6	9	10	5	−1	−1
1	2	8	7	0	−1	6	1	4	7	6	2	−1	−1	1	6	10	15	5	−1	4
1	2	9	10	0	−1	9	1	4	8	7	2	−1	2	1	7	8	7	6	−1	−6
1	2	10	15	0	−1	14	1	4	9	10	2	−1	5	1	7	9	10	6	−1	−3
1	3	4	2	1	−1	−1	1	4	10	15	2	−1	10	1	7	10	15	6	−1	−2
1	3	5	3	1	−1	0	1	5	6	5	3	−1	−2	1	8	9	10	7	−1	−5
1	3	6	5	1	−1	2	1	5	7	6	3	−1	−1	1	8	10	15	7	−1	0
1	3	7	6	1	−1	3	1	5	8	7	3	−1	0	1	9	10	15	10	−1	−6

Окончание

i	j	k	x_k	x_j	x_i	x_{kji}	i	j	k	x_k	x_j	x_i	x_{kji}	i	j	k	x_k	x_j	x_i	x_{kji}
2	3	4	2	1	0	0	3	4	5	3	2	1	0	4	6	9	10	5	2	2
2	3	5	3	1	0	1	3	4	6	5	2	1	2	4	6	10	15	5	2	7
2	3	6	5	1	0	3	3	4	7	6	2	1	3	4	7	8	7	6	2	−3
2	3	7	6	1	0	4	3	4	8	7	2	1	4	4	7	9	10	6	2	0
2	3	8	7	1	0	5	3	4	9	10	2	1	7	4	7	10	15	6	2	5
2	3	9	10	1	0	8	3	4	10	15	2	1	12	4	8	9	10	7	2	−2
2	3	10	15	1	0	13	3	5	6	5	3	1	0	4	8	10	15	7	2	3
2	4	5	3	2	0	−1	3	5	7	6	3	1	1	4	9	10	15	10	2	−3
2	4	6	15	2	0	1	3	5	8	7	3	1	2	5	6	7	6	5	3	−1
2	4	7	6	2	0	2	3	5	9	10	3	1	5	5	6	8	7	5	3	0
2	4	8	7	2	0	3	3	5	10	15	3	1	10	5	6	9	10	5	3	3
2	4	9	10	2	0	6	3	6	7	6	5	1	−3	5	6	10	15	5	3	8
2	4	10	15	2	0	11	3	6	8	7	5	1	−2	5	7	8	7	6	3	−2
2	5	6	5	3	0	−1	3	6	9	10	5	1	1	5	7	9	10	6	3	1
2	5	7	6	3	0	0	3	6	10	15	5	1	6	5	7	10	15	6	3	6
2	5	8	7	3	0	1	3	7	8	7	6	1	−4	5	8	9	10	7	3	−1
2	5	9	10	3	0	4	3	7	9	10	6	1	−1	5	8	10	15	7	3	4
2	5	10	15	3	0	9	3	7	10	15	6	1	4	5	9	10	15	10	3	−2
2	6	7	6	5	0	−4	3	8	9	10	7	1	−3	5	7	8	7	6	5	0
2	6	8	7	5	0	−3	3	8	10	15	7	1	2	6	7	9	10	6	5	3
2	6	9	10	5	0	0	3	9	10	15	10	1	−4	6	7	10	15	6	5	8
2	6	10	15	5	0	5	4	5	6	5	3	2	1	6	8	9	10	7	5	1
2	7	8	7	6	0	−5	4	5	7	6	3	2	2	6	8	10	15	7	5	6
2	7	9	10	6	0	−2	4	5	8	7	3	2	3	6	9	10	15	10	5	0
2	7	10	15	6	0	3	4	5	9	10	3	2	6	7	8	9	10	7	6	2
2	8	9	10	7	0	−4	4	5	10	15	3	2	11	7	8	10	15	7	6	7
2	8	10	15	7	0	1	4	6	7	6	5	2	−2	7	9	10	15	10	6	1
2	9	10	15	10	0	−5	4	6	8	7	5	2	−1	8	9	10	15	10	7	2

Имеем

$$\sum_{1\leqslant i<j<k}(x_k-2x_j+x_i)=234;\quad U_{1n}=\frac{1}{C_{10}^3}\cdot 234=\frac{234}{120}=1{,}95;\quad T_{1n}=\frac{1{,}95}{4{,}9396}=0{,}3948.$$

Используем аппроксимации:

$$\mathbf{M}(T_{1n})=0;\quad \mathbf{D}(T_{1n})=\frac{1}{120}\cdot(1{,}03804+0{,}69714\cdot 7+0{,}0890805\cdot 7\cdot 6)=0{,}080495$$
$$(\sqrt{\mathbf{D}(T_{1n})}=0{,}2837).$$

Для $u_{0,95}=1{,}645$ получаем $T_{1n}(0{,}05)\approx 0{,}2837\cdot 1{,}645=0{,}466$.

Так как $T_{1n}=0{,}3948<T_{1n}(0{,}05)=0{,}466$, нормальность распределения вероятностей случайных величин не отклоняется. К аналогичному выводу (читатель может сам убедиться) приводит и использование статистики T_{2n}.

3.2.2.8. Критерий нормальности Оя

Критерий, аналогичный рассмотренному выше (см. раздел 3.2.2.7), использующий анализ комбинаций порядковых статистик, рассмотрен в работах Оя [258–260].

Статистиками критерия являются

$$T_1=\frac{1}{C_n^3}\sum_{1\leqslant i<j<k\leqslant n}\frac{x_k-x_j}{x_k-x_i};\quad T_2=\frac{1}{C_n^4}\sum_{1\leqslant i<j<k<l\leqslant n}\frac{x_k-x_j}{x_l-x_i}.$$

Статистика T_1 эффективна против альтернатив, отличающихся от нормального распределения коэффициентом асимметрии, а T_2 — коэффициентом эксцесса. При справедливости нулевой гипотезы $M(T_1) = 0{,}5$ и $M(T_2) = 0{,}298746$ [259], дисперсии приведены в табл. 73.

Таблица 73

Дисперсии статистик Оя [259]

n	$\mathbf{D}(T_1)$	$\mathbf{D}(T_2)$	n	$\mathbf{D}(T_1)$	$\mathbf{D}(T_2)$
5	0,01512	0,01549	10	0,00358	0,00196
6	0,01045	0,00800	15	0,00218	0,00098
7	0,00740	0,00473	20	0,00144	0,00058
8	0,00575	0,00350	30	0,00089	0,00032
9	0,00457	0,00248	∞	$0{,}0214/n$	$0{,}0026/n$

Статистики

$$\tilde{T} = \frac{T_1 - 0{,}5}{\sqrt{\mathbf{D}(T_1)}} \quad \text{и} \quad \tilde{T}_2 = \frac{T_2 - 0{,}298746}{\sqrt{\mathbf{D}(T_2)}}$$

распределены приближенно нормально, и их критические значения могут быть вычислены по формулам

$$\tilde{T}_1(\alpha) = 0{,}5 + \sqrt{\mathbf{D}(T_1)}u_\alpha \quad \text{и} \quad \tilde{T}_2(\alpha) = 0{,}298746 + \sqrt{\mathbf{D}(T_2)}u_\alpha,$$

где u_α — α-квантиль стандартного нормального распределения.

Учитывая независимость статистик $\tilde{T}_1, \tilde{T}_2$, можно рекомендовать комбинированный критерий

$$\chi^2 = \tilde{T}_1^2 + \tilde{T}_2^2,$$

имеющий распределение χ^2 с $f = 2$ степенями свободы.

Более просты в вычислительном отношении модифицированные критерии [260]

$$T_1' = \sum_{1 \leqslant i < j \leqslant n} a_{ij} \log(x_j - x_i) \quad \text{и} \quad T_2' = \sum_{1 \leqslant i < j \leqslant n} b_{ij} \log(x_j - x_i),$$

где $a_{ij} = \frac{1}{C_n^3}(i + j - n - 1)$; $b_{ij} = \frac{1}{C_n^4}\left[2(n - j)(i - 1) - C_{n-j}^2 - C_{i-1}^2\right]$.

При справедливости нулевой гипотезы

$$\mathbf{M}(T_1') = 0; \quad \mathbf{M}(T_2') = 0{,}4523;$$

$$\mathbf{D}(T_1') = \frac{1}{C_n^3}(0{,}07478C_3^1 + 0{,}03963C_3^2C_{n-3}^1 + 2{,}8979) = \\ = \frac{1}{C_n^3}\left[0{,}11217(n - 4)(n - 3)).0{,}11889(n - 3) + 2{,}8979\right];$$

$$\mathbf{D}(T_2') = \frac{1}{C_n^4}(0{,}1311C_4^1C_{n-4}^3 + 0{,}0145C_4^2C_{n-4}^2 + 1{,}3355C_4^3C_{n-4}^1 + 8{,}8552) = \\ = \frac{1}{C_n^4}\left[0{,}0874(n - 6)(n - 5)(n - 4) + 0{,}0435(n - 5)(n - 4) + 5{,}342(n - 4) + 8{,}8552\right].$$

Как и прежде, $T_1'(\alpha) = \sqrt{\mathbf{D}(T_1')}\,u_\alpha$, $T_2'(\alpha) = 0{,}4523 + \sqrt{\mathbf{D}(T_2')}\,u_\alpha$.

Нормальная аппроксимация статистик T_1' и T_2' несколько хуже, чем T_1 и T_2. По-прежнему можно использовать критерий

$$\chi^2 = (\tilde{T}_1)^2 + (\tilde{T}_2)^2 = \frac{\left[T_1' - \mathbf{M}(T_1')\right]^2}{\mathbf{D}(T_1')} + \frac{\left[T_2' + \mathbf{M}(T_2')\right]^2}{\mathbf{D}(T_2')}.$$

Задача 114. *В условиях задачи* 107 *проверить гипотезу нормальности распределения вероятностей критерием Оя на уровне значимости* $\alpha = 0{,}05$ (*доверительная вероятность* 0,95).

Воспользовавшись данными таблицы задачи 113, получим ряд значений $x_{kji} = \dfrac{x_k - x_j}{x_k - x_i}$. Результаты расчетов сведены в таблицу:

i	j	k	x_{kji}	i	j	k	x_{kji}	i	j	k	x_{kji}	i	j	k	x_{kji}
1	2	3	1/2	1	7	8	1/8	2	7	10	9/15	4	6	7	1/4
1	2	4	2/3	1	7	9	4/11	2	8	9	3/10	4	6	8	2/5
1	2	5	3/4	1	7	10	9/16	2	8	10	8/15	4	6	9	5/8
1	2	6	5/6	1	8	9	3/11	2	9	10	5/15	4	6	10	10/13
1	2	7	6/7	1	8	10	8/16	3	4	5	1/2	4	7	8	1/5
1	2	8	7/8	1	9	10	5/16	3	4	6	3/4	4	7	9	4/8
1	2	9	10/11	2	3	4	1/2	3	4	7	4/5	4	7	10	9/13
1	2	10	15/16	2	3	5	2/3	3	4	8	5/6	4	8	9	3/8
1	3	4	1/3	2	3	6	4/5	3	4	9	8/9	4	8	10	8/13
1	3	5	2/4	2	3	7	5/6	3	4	10	13/14	4	9	10	5/13
1	3	6	4/6	2	3	8	6/7	3	5	6	2/4	5	6	7	1/3
1	3	7	5/7	2	3	9	9/10	3	5	7	3/5	5	6	8	2/4
1	3	8	6/8	2	3	10	14/15	3	5	8	4/6	5	6	9	5/7
1	3	9	9/11	2	4	5	1/3	3	5	9	7/9	5	6	10	10/12
1	3	10	14/16	2	4	6	3/5	3	5	10	12/14	5	7	8	1/4
1	4	5	1/4	2	4	7	4/6	3	6	7	1/5	5	7	9	4/7
1	4	6	3/6	2	4	8	5/7	3	6	8	2/6	5	7	10	9/12
1	4	7	4/7	2	4	9	8/10	3	6	9	5/9	5	8	9	3/7
1	4	8	5/8	2	4	10	13/15	3	6	10	10/14	5	8	10	8/12
1	4	9	8/11	2	5	6	2/5	3	7	8	1/6	5	9	10	5/12
1	4	10	13/16	2	5	7	3/6	3	7	9	4/9	6	7	8	1/2
1	5	6	2/6	2	5	8	4/7	3	7	10	10/14	6	7	9	4/5
1	5	7	3/7	2	5	9	7/10	3	8	9	3/9	6	7	10	9/10
1	5	8	4/8	2	5	10	12/15	3	8	10	8/14	6	8	9	3/5
1	5	9	7/11	2	6	7	1/6	3	9	10	5/14	6	8	10	8/10
1	5	10	12/16	2	6	8	2/7	4	5	6	2/3	6	9	10	5/10
1	6	7	1/7	2	6	9	5/10	4	5	7	3/4	7	8	9	3/4
1	6	8	2/8	2	6	10	10/15	4	5	8	4/5	7	8	10	8/9
1	6	9	5/11	2	7	8	1/7	4	5	9	7/8	7	9	10	5/9
1	6	10	10/16	2	7	9	4/10	4	5	10	12/13	8	9	10	5/8

Из таблицы находим

$$\sum_{1 \leqslant i < j < k \leqslant n} \frac{x_k - x_j}{x_k - x_1} = 70{,}6029 \quad \text{и} \quad T_1 = \frac{1}{C_n^3} \cdot 70{,}6029 = 0{,}58836.$$

Далее $\mathbf{M}(T_1) = 0{,}5$ и $\mathbf{D}(T_1) = 0{,}00358$ (см. табл. 73).

Окончательно ($u_{0,95} = 1{,}645$): $T_1(0{,}95) = 0{,}5 + \sqrt{0{,}00358} \cdot 1{,}645 = 0{,}5984$.

Так как $T_1 = 0{,}58836 < T_1(0{,}95) = 0{,}5984$, гипотеза нормальности распределения случайных величин не отклоняется. К аналогичному результату приводят другие критерии (читатель может это проверить самостоятельно).

3.2.2.9. Критерий среднего абсолютного отклонения (критерий Гири)

Гири [261–263] рассмотрел критерий нормальности распределения случайных величин, основанный на статистике

$$d = \frac{1}{ns}\sum_{i=1}^{n}|x_i - \bar{x}|,$$

являющейся отношением среднего абсолютного отклонения $\frac{1}{n}\sum\limits_{i=1}^{n}|x_i - \bar{x}|$ к выборочному стандартному отклонению

$$s = \left\{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{1}{2}}.$$

Статистика Гири распределена асимптотически нормально при $n \geqslant 50$ со средним

$$\mathbf{M}(d) = \frac{2}{\sqrt{\pi(n-1)}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \approx \sqrt{\frac{2}{\pi}}\left(1 + \frac{2}{8n-9}\right) \approx 0{,}79788\frac{n-0{,}875}{n-1{,}125}$$

и дисперсией

$$\mathbf{D}(d) = \frac{1}{n}\left\{1 + \frac{2}{\pi}\left[\sqrt{n(n-2)} + \arcsin\left(\frac{1}{n-1}\right)\right]\right\} - \frac{n-1}{\pi}\left[\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\right]^2 =$$

$$= \frac{1}{n}\left[\left(1 - \frac{3}{\pi}\right) - \frac{1}{4n\pi}\right] \approx \frac{1}{n}\left(0{,}04507 - \frac{0{,}0796}{n}\right).$$

Более точные аппроксимации приведены в [262]:

$$\mathbf{M}(d) = 0{,}797885 + \frac{0{,}199471}{n} + \frac{0{,}024934}{n^2} - \frac{0{,}031168}{n^3} - \frac{0{,}008182}{n^4};$$

$$\mathbf{D}(d) = \frac{0{,}045070}{n} - \frac{0{,}124648}{n^2} + \frac{0{,}084859}{n^3} + \frac{0{,}006323}{n^4}.$$

Для практических расчетов рекомендуется формула $\sum\limits_{i=1}^{n}|x_i - \bar{x}| = 2(\sum x' - \bar{x}n')$, где $\sum x'$ — сумма значений x-ов, превышающих $\bar{x}$; n' — количество значений x-ов, превышающих $\bar{x}$.

Критерий двусторонний. При $n \geqslant 50$ вычисляются квантили

$$d\left(\frac{\alpha}{2}\right) = \mathbf{M}(d) + \sqrt{\mathbf{D}(d)}\,u_{\frac{\alpha}{2}}; \quad d\left(1 - \frac{\alpha}{2}\right) = \mathbf{M}(d) + \sqrt{\mathbf{D}(d)}\,u_{1-\frac{\alpha}{2}},$$

где u_α — α-квантиль стандартного нормального распределения; α — уровень значимости.

Если $d\left(\frac{\alpha}{2}\right) \leqslant d \leqslant d\left(1 - \frac{\alpha}{2}\right)$, то гипотеза нормальности распределения случайных величин не отклоняется.

Некоторые критические значения d-статистики приведены в табл. 74 (в таблице приняты обозначения $d_1(\alpha)$ и $d_2(\alpha)$, соответствующие неравенству $d_1(\alpha) \leqslant d \leqslant d_2(\alpha)$, которые удовлетворяются при справедливости нулевой гипотезы).

Таблица 74

Критические значения $d_1(\alpha)$ и $d_2(\alpha)$ критерия Гири

(α — уровень значимости) [262]

n	$\alpha = 0{,}01$		$\alpha = 0{,}05$		$\alpha = 0{,}10$	
	d_1	d_2	d_1	d_2	d_1	d_2
11	0,1175	0,9359	0,7153	0,9073	0,7409	0,8899
16	0,6829	0,9137	0,7236	0,8884	0,7452	0,8733
21	0,6950	0,9001	0,7304	0,8768	0,7495	0,8631
26	0,7040	0,8901	0,7360	0,8686	0,7530	0,8570
31	0,7110	0,8827	0,7404	0,8625	0,7559	0,8511
36	0,7167	0,8769	0,7440	0,8578	0,7583	0,8468
41	0,7215	0,8722	0,7470	0,8540	0,7604	0,8436
46	0,7256	0,8682	0,7496	0,8508	0,7621	0,8409
51	0,7291	0,8648	0,7518	0,8481	0,7636	0,8385
61	0,7347	0,8592	0,7554	0,8434	0,7662	0,8349
71	0,7393	0,8549	0,7583	0,8403	0,7683	0,8321
81	0,7430	0,8515	0,7607	0,8376	0,7700	0,8298
91	0,7460	0,8484	0,7626	0,8353	0,7714	0,8279
101	0,7487	0,8460	0,7644	0,8344	0,7726	0,8264

Задача 115. *В условиях задачи* 107 *проверить гипотезу нормальности распределения случайных величин критерием Гири на уровне значимости* $\alpha = 0{,}05$.

Имеем

$$\bar{x} = 4{,}8; \quad s = \left\{ \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}^{\frac{1}{2}} = 4{,}686;$$

$$\sum_{i=1}^{n} |x_i - \bar{x}| = |-1 - 4{,}8| + |0 - 4{,}8| + |1 - 4{,}8| + \ldots + |15 - 4{,}8| = 38.$$

По упрощенной формуле ($n' = 5$ значений превышают $\bar{x} = 4{,}8$; их сумма равна $\sum x' = 5 + 6 + 7 + 10 + 15 = 43$) получаем $\sum |x_i - \bar{x}| = 2 \cdot (43 - 4{,}8 \cdot 5) = 38$.

Статистика критерия равна $d = \dfrac{1}{10 \cdot 1{,}686} \cdot 38 = 0{,}8109$.

Далее

$$\mathbf{M}(d) = 0{,}79788 \cdot \frac{10 - 0{,}875}{10 - 1{,}125} = 0{,}8203;$$

$$\mathbf{D}(d) = \frac{1}{10} \cdot \left(0{,}04507 - \frac{0{,}0796}{10} \right) = 3{,}711 \cdot 10^{-3} \quad (\sqrt{D(d)} = 0{,}0609).$$

Из табл. 74 находим для $n = 10$ и $\alpha = 0{,}05$: $d_1 \approx 0{,}715$ и $d_2 \approx 0{,}907$.

Так как $0{,}715 < d = 0{,}8109 < 0{,}907$, нулевая гипотеза нормальности не отклоняется.

Воспользуемся нормальным приближением ($u_{0,95} = 1{,}645$ и $u_{0,05} = -1{,}645$),

$$d_1(0{,}05) = M(d) = +\sqrt{\mathbf{D}(d)} \cdot U_{0,95} = 0{,}8203 - 1{,}645 \cdot 0{,}0609 = 0{,}720;$$

$$d_2(0{,}05) = 0{,}8202 + 1{,}645 \cdot 0{,}0609 = 0{,}920.$$

Так как $0{,}720 < d = 0{,}8109 < 0{,}920$, нулевая гипотеза нормальности не отклоняется.

3.2.2.10. Критерий Дэвида–Хартли–Пирсона

В [264] предложен критерий нормальности распределения вероятностей случайной величины, основанный на распределении отношения размаха к стандартному отклонению.

Статистика критерия имеет вид

$$U = \frac{R}{s},$$

где $R = x_{\max} - x_{\min}$ (или $(x_n - x_1)$ для упорядоченного по возрастанию ряда выборочных значений) — размах выборки; s — стандартное отклонение.

Таблица 75

Критические границы $U_1(\alpha)$ и $U_2(\alpha)$ критерия Дэвида–Хартли–Пирсона

(α — уровень значимости) [9]

n	$\alpha = 0{,}01$		$\alpha = 0{,}05$		$\alpha = 0{,}10$	
	U_1	U_2	U_1	U_2	U_1	U_2
3	1,737	2,000	1,758	1,999	1,782	1,997
4	1,870	2,445	1,980	2,429	2,040	2,409
5	2,020	2,803	2,150	2,753	2,220	2,712
6	2,150	3,095	2,200	3,012	2,370	2,949
7	2,260	3,338	2,400	3,222	2,490	3,143
8	2,350	3,543	2,500	3,399	2,590	3,308
9	2,440	3,720	2,590	3,552	2,680	3,449
10	2,510	3,875	2,670	3,685	2,760	3,570
11	2,580	4,012	2,740	3,800	2,840	3,680
12	2,640	4,134	2,800	3,910	2,900	3,780
13	2,700	4,244	2,860	4,000	2,960	3,870
14	2,750	4,340	2,920	4,090	3,020	3,950
15	2,800	4,440	2,970	4,170	3,070	4,020
16	2,840	4,520	3,010	4,240	3,120	4,090
17	2,880	4,600	3,060	4,310	3,170	4,150
18	2,920	4,670	3,100	4,370	3,210	4,210
19	2,960	4,740	3,140	4,430	3,250	4,270
20	2,990	4,800	3,180	4,490	3,290	4,320
25	3,150	5,060	3,340	4,710	3,450	4,530
30	3,270	5,260	3,470	4,890	3,590	4,700
35	3,380	5,420	3,580	5,040	3,700	4,840
40	3,470	5,560	3,670	5,160	3,790	4,960
45	3,550	5,670	3,750	5,260	3,880	5,060
50	3,620	5,770	3,830	5,350	3,950	5,140
55	3,690	5,860	3,900	5,430	4,020	5,220
60	3,750	5,940	3,960	5,510	4,080	5,290
65	3,800	6,010	4,010	5,570	4,140	5,350
70	3,850	6,070	4,060	5,630	4,190	5,410
75	3,900	6,130	4,130	5,680	4,240	5,460
80	3,940	6,180	4,150	5,730	4,280	5,510
85	3,990	6,230	4,200	5,780	4,330	5,560
90	4,020	6,270	4,240	5,820	4,360	5,600
95	4,060	6,320	4,270	5,860	4,400	5,640
100	4,100	6,360	4,310	5,900	4,440	5,680

Гипотеза нормальности принимается, если $U_1(\alpha) < U < U(\alpha)$ (α — уровень значимости). Наиболее полные таблицы критических значений $U_1(\alpha)$ приведены в [9, 256] (в табл. 75 приведены некоторые значения $U_1(\alpha)$ и $U_2(\alpha)$).

При $n \to \infty$ имеют место соотношения [266]

$$2\sqrt{\frac{n-1}{n}} \leqslant \frac{R}{s} \leqslant \sqrt{2(n-1)} \quad \text{при } n = 2, 4, \ldots, 2k;$$

$$2\sqrt{\frac{n}{n+1}} \leqslant \frac{R}{s} \leqslant \sqrt{2(n-1)} \quad \text{при } n = 1, 3, \ldots, 2k-1.$$

Задача 116. *В условиях задачи* 107 *проверить гипотезу нормальности распределения вероятностей случайных величин критерием Дэвида–Хартли–Пирсона при* $\alpha = 0{,}05$.

Имеем $R = x_{\max} - x_{\min} = 15 - (-1) = 16$; $s = 4{,}686$; $U = \dfrac{R}{s} = \dfrac{16}{4{,}686} = 3{,}414$.

Из табл. 75 для $n = 10$ и $\alpha = 0{,}05$ имеем $U_1 = 2{,}67$ и $U_2 = 3{,}522$.

Так как $2{,}67 < U = 3{,}414 < 3{,}552$, гипотеза нормальности распределения случайных величин не отклоняется на уровне значимости 0,05.

3.2.2.11. Комбинированный критерий Шпигельхальтера

В [267, 268] рассмотрен комбинированный критерий нормальности против симметричных альтернатив с неизвестными средним и дисперсией. Критерий основан на комбинации двух критериев — критерия Гири (см. раздел 3.2.2.9) и критерия Дэвида–Хартли–Пирсона (см. раздел 3.2.2.10).

Статистика критерия имеет вид

$$T' = \left\{(C_n U)^{-(n-1)} + g^{-(n-1)}\right\}^{\frac{1}{n-1}},$$

где $C_n = \dfrac{1}{2n}(n!)^{\frac{1}{n-1}}$; $U = \dfrac{R}{s}$; $g = \dfrac{\sum\limits_{i=1}^{n}|x_i - \bar{x}|}{s\sqrt{n(n-1)}}$.

Нулевая гипотеза нормальности испытываемого распределения принимается, если $T' < T'(\alpha)$. Критические значения $T'(\alpha)$ на уровнях значимости $\alpha = 0{,}05$ и $0{,}10$, а также значения коэффициента C_n приведены в табл. 76.

Таблица 76

Критические значения $T'(\alpha)$ критерия Шпигельхальтера [267]

n	α		C_n
	0,05	0,10	
5	1,532	1,512	0,3310
10	1,453	1,417	0,2678
15	1,423	1,387	0,2445
20	1,403	1,369	0,2321
50	1,337	1,317	0,2070
100	1,308	1,294	0,1971

Критерий обладает хорошей мощностью против симметричных альтернатив.

Задача 117. *В условиях задачи* 107 *проверить гипотезу нормальности распределения вероятностей случайной величины критерием Шпигельхальтера при* $\alpha = 0{,}05$.

Имеем $R = 15 - (-1) = 16$ и $s = 4{,}686$.

Из табл. 76 находим $C_n = 0{,}2678$ и вычисляем

$$U = \frac{R}{s}; \quad g = \frac{38}{4{,}686 \cdot \sqrt{10 \cdot 9}} = 0{,}8548; \quad T' = \left\{(0{,}2678 \cdot 3{,}414)^{-9} + 0{,}8548^{-9}\right\}^{\frac{1}{9}} = 1{,}228.$$

Из табл. 76 для $\alpha = 0{,}05$ имеем $T'(0{,}05) = 1{,}453$.

Так как $T' = 1{,}228 < T'(0{,}05) = 1{,}453$, гипотеза нормальности не отклоняется.

3.2.2.12. Критерий нормальности Саркади

Саркади [269] предложил критерий нормальности, более совершенный, чем другие. Предположим, имеется выборка значений случайной величины $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$. Критерий Саркади строится следующим образом.

Образуем статистику

$$z_j = y_j \left\{ \frac{1}{n-j-1} \left(y_{i+1}^2 + \ldots + y_{n-1}^2\right) \right\}^{-\frac{1}{2}} \quad (j = 1, 2, \ldots, n-2),$$

где

$$y_j = x_j - \frac{1}{n + \sqrt{n}} \sum_{i=1}^{n} x_i - \frac{1}{\sqrt{n} + 1} x_m \quad (j = m, \ldots, n-1);$$

m — произвольное фиксированное число ($1 \leqslant m \leqslant n$).

Если сложная гипотеза нормальности верна, то переменные z_j взаимно независимы и имеют t-распределение (см. раздел 1.1.9) с $f = n - j - 1$ степенями свободы. Тогда случайные величины

$$r_j = F_{n-j-1}(z_j), \quad j = 1, 2, \ldots, n-2,$$

распределены равномерно на интервале $[0, 1]$ ($F_f(z)$ — функция распределения Стьюдента с f степенями свободы).

Таким образом, задача сводится к проверке равномерности распределения величин r_j по одному из критериев согласия (см. раздел 3.4). Если проверяется только факт нормальности распределения, то следует выбирать $m = n$ или $m = 1$, в ином случае (например, тренд в среднем, изменение дисперсии в альтернативе) используются другие значения m.

В [270] предлагается использовать для проверки равномерности распределения r_j „гладкий критерий" Неймана, статистика которого имеет вид

$$P_4^2 = \frac{1}{n-2} \sum_{i=1}^{4} \sum_{j=1}^{n-2} [\varphi_i(r_j)]^2,$$

где

$$\varphi_1(y) = \sqrt{12}\left(y - \frac{1}{2}\right); \quad \varphi_2(y) = \sqrt{5}\left[61\left(y - \frac{1}{2}\right)^2 - \frac{1}{2}\right];$$

$$\varphi_3(y) = \sqrt{7}\left[20\left(y - \frac{1}{2}\right)^3 - 3\left(y - \frac{1}{2}\right)\right]; \quad \varphi_4(y) = 210\left(y - \frac{1}{2}\right)^4 - 45\left(y - \frac{1}{2}\right)^2 + \frac{9}{8}$$

— полиномы Лежандра, ортогональные на отрезке $[0, 1]$.

Распределение статистики P_4^2 близко к распределению χ^2 с $f = 4$ степенями свободы. Следовательно, если $P_4^2 < \chi_f^2(\alpha)$, то на уровне значимости α гипотеза нормальности не отклоняется.

Функция распределения Стьюдента достаточно полно табулирована в [25, табл. 3.1а, с. 236]. В связи с большим объемом этой таблицы она здесь не

приводится, читатель может легко найти ее в [25]. При необходимости можно воспользоваться аппроксимациями раздела 1.1.9.

Задача 118. *Проверить гипотезу нормальности распределения случайных величин критерием Саркади при* $\alpha = 0{,}05$.

Выбираем $m = n = 10$ ($\sum x_i = 48$) и вычисляем для $j = 1, 2, \ldots, 9$

$$y_j = x_j - \frac{1}{n+\sqrt{n}} \cdot \sum x_i - \frac{1}{\sqrt{n}+1} \cdot x_{10} = x_j - 7{,}25031;$$

$$y_1 = -8{,}25031; \quad y_2 = -7{,}25031; \quad y_3 = -6{,}02531; \quad y_4 = -5{,}02531; \quad y_5 = -4{,}02531;$$
$$y_6 = -2{,}2531; \quad y_7 = -1{,}2531; \quad y_8 = -0{,}2531; \quad y_9 = 2{,}74969.$$

Далее находим

$$z_1 = y_1 \cdot \left\{\frac{1}{8} \cdot \left(y_2^2 + \ldots + y_9^2\right)\right\}^{-\frac{1}{2}} = -1{,}89577;$$

$$z_2 = y_2 \cdot \left\{\frac{1}{7} \cdot \left(y_3^2 + \ldots + y_9^2\right)\right\}^{-\frac{1}{2}} = -1{,}92842;$$

$$z_3 = -1{,}97848; \quad z_4 = -2{,}0652; \quad z_5 = -2{,}2519;$$
$$z_6 = -1{,}2859; \quad z_7 = -0{,}6404; \quad z_8 = -0{,}09103.$$

Переменные r_j находим, пользуясь таблицей [25, табл. 3.1а, с. 236] функции t-распределения Стьюдента

$$r_1 = F_8(-1{,}896) = 1 - F_8(1{,}896) \approx 1 - 0{,}956 = 0{,}044; \quad r_2 = F_7(-1{,}928) = 0{,}049;$$
$$r_3 = F_6(-1{,}978) = 0{,}046; \quad r_4 = F_5(-2{,}065) = 0{,}048; \quad r_5 = F_4(-2{,}252) = 0{,}044;$$
$$r_6 = F_3(-1{,}286) = 0{,}150; \quad r_7 = F_2(-0{,}540) = 0{,}317; \quad r_8 = F_1(-0{,}091) = 0{,}469.$$

Теперь вычислим статистику „гладкого критерия“ Неймана.

Находим

$$\varphi_1(r_1) = \sqrt{12} \cdot (0{,}044 - 0{,}5) = -1{,}5796; \quad \varphi_2(r_1) = \sqrt{5} \cdot \left[(0{,}044 - 0{,}5)^2 - 0{,}5\right] = 1{,}6717;$$

$$\varphi_3(r_1) = \sqrt{7} \cdot \left[20 \cdot (0{,}044 - 0{,}5)^3 - 3 \cdot (0{,}044 - 0{,}5)\right] = -1{,}3979;$$

$$\varphi_4(r_1) = 210 \cdot (0{,}044 - 0{,}5)^4 - 45 \cdot (0{,}044 - 0{,}5)^2 + \frac{9}{8} = 0{,}8477;$$

$$\varphi_1(r_2) = -1{,}5623; \quad \varphi_2(r_2) = 1{,}61087; \quad \varphi_3(r_2) = -1{,}2744; \quad \varphi_4(r_2) = 0{,}6601;$$
$$\varphi_1(r_3) = -1{,}5727; \quad \varphi_2(r_3) = 1{,}6473; \quad \varphi_3(r_3) = -1{,}3480; \quad \varphi_4(r_3) = 0{,}7714;$$
$$\varphi_1(r_4) = -1{,}5638; \quad \varphi_2(r_4) = 1{,}6230; \quad \varphi_3(r_4) = -1{,}2988; \quad \varphi_4(r_4) = 0{,}6967;$$
$$\varphi_1(r_5) = -1{,}3979; \quad \varphi_2(r_5) = 1{,}6717; \quad \varphi_3(r_5) = -1{,}3979; \quad \varphi_4(r_5) = 0{,}8477;$$
$$\varphi_1(r_6) = -1{,}2124; \quad \varphi_2(r_6) = 0{,}5255; \quad \varphi_3(r_6) = 0{,}5093; \quad \varphi_4(r_6) = -1{,}2362;$$
$$\varphi_1(r_7) = -0{,}6339; \quad \varphi_2(r_7) = -0{,}6687; \quad \varphi_3(r_7) = 1{,}1282; \quad \varphi_4(r_7) = -0{,}1465;$$
$$\varphi_1(r_8) = -0{,}1074; \quad \varphi_2(r_8) = -1{,}1051; \quad \varphi_3(r_8) = 0{,}2445; \quad \varphi_4(r_8) = 1{,}0819.$$

Окончательно получаем

$$\sum_{j=1}^{8} [\varphi_1(r_j)]^2 = 14{,}2394; \quad \sum_{j=1}^{8} [\varphi_2(r_j)]^2 = 12{,}68176;$$

$$\sum_{j=1}^{8} [\varphi_3(r_j)]^2 = 10{,}6257; \quad \sum_{j=1}^{8} [\varphi_4(r_j)]^2 = 5{,}8756; \quad P_4^2 = \frac{1}{8} \sum_{i=1}^{4} \sum_{j=1}^{8} [\varphi_i(r_j)]^2 = 5{,}4278.$$

Критические значения $\chi_4^2(1-\alpha)$ находим из табл. 55: $\chi_4^2(0{,}95) = 9{,}488$.

Так как $P_4^2 = 5{,}4278 < \chi_4^2(0{,}95) = 9{,}488$, гипотеза нормальности распределения случайных величин не отклоняется.

3.2.2.13. Критерий нормальности Лина–Мудхолкара

В [271–273] показано, что если $x_1, \ldots, x_n$ $(n \geqslant 2)$ — выборка из распределения с функцией $F(x)$, то $\bar{x}$ и $s^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}\left(x_i - \bar{x}^2\right)$ независимы тогда и только тогда, когда $F(x)$ — функция распределения нормального закона. Эта характеризация использована в [274] для построения критерия проверки нормальности распределения против асимметричных альтернатив. Критерий позволяет проверить сложную гипотезу нормальности распределения с неизвестными средним и (или) дисперсией.

Применение этого критерия не требует ни упорядочения, ни преобразования переменных, ни выборочной оценки параметров распределения. Распределение статистики этого критерия при H_0 очень близко к нормальному в малых выборках (тогда как в других критериях это имеет место, как правило, только в асимптотике).

Критерий основан на анализе n средних и дисперсий, рассчитанных по n выборкам, образуемым исключением каждый раз одного наблюдения. Хотя полученные пары значений $(\bar{x}_i, s_i^2)$ не будут независимыми, их можно использовать для построения критерия.

Одно из распределений — распределение s^2 не является нормальным, а критерий предполагает использование коэффициента корреляции в нормальной двумерной совокупности. Поэтому для дисперсии применяется нормализующее преобразование Вилсона–Хилферти

$$y_i = \left\{\frac{1}{n}\left[\sum_{j\neq i} x_j^2 - \frac{1}{n-1}\left(\sum_{j\neq i} x_j\right)^2\right]\right\}^{\frac{1}{3}}, \quad i = 1, \ldots, n.$$

Величину y_i можно рассматривать как нормально распределенную случайную величину.

Таким образом, исследуется корреляция между величинами $\bar{x}$ и y_i, оцениваемая коэффициентом корреляции

$$r = \frac{\sum\limits_{i=1}^{n}(\bar{x}_i - \bar{\bar{x}})(y_i - \bar{y})}{\left\{\sum\limits_{i=1}^{n}(\bar{x}_i - \bar{\bar{x}})^2 \sum\limits_{i=1}^{n}(y_i - \bar{y})^2\right\}^{\frac{1}{2}}}.$$

Для построения статистики критерия используется нормализующее преобразование $z = \dfrac{1}{2}\ln\dfrac{1+r}{1-r}$. Дисперсия $\sigma^2(z)$ и коэффициент эксцесса $\alpha_4(z)$ величины z равны

$$\sigma^2(z) = \frac{3}{n} - \frac{7{,}324}{n^2} + \frac{53{,}005}{n^3}; \quad \alpha_4(z) = -\frac{11{,}7}{n} + \frac{55{,}06}{n^2}.$$

Критические значения $z(\alpha)$ могут быть выражены через квантили стандартного нормального распределения u_α следующим образом:

$$z(\alpha) = \sigma(z)\left\{u_{\frac{1+\alpha}{2}} + \frac{1}{24}\left(u^3_{\frac{1+\alpha}{2}} - 3u_{\frac{1+\alpha}{2}}\right)\alpha_4(z)\right\},$$

где α — уровень значимости.

Если $\bar{x}$ и s^2 — среднее и дисперсия полной выборки, то $\bar{x}_i$, s_i^2 — среднее и дисперсия по всем наблюдениям, исключая i-е:

$$\bar{x}_i = \frac{1}{n-1}(n\bar{x} - x_i); \quad s_i^2 = \frac{1}{n-2}\left[(n-1)\,s^2 - \frac{n}{n-1}(x_i - \bar{x})^2\right].$$

Программа расчета и некоторые практические рекомендации по применению этого критерия рассмотрены в [275, 276].

В конечном итоге статистикой критерия является коэффициент корреляции между $\bar{x}_i$ и $\left(s_i^2\right)^{\frac{1}{3}}$. Если $|z| < z\left(\dfrac{1+\alpha}{2}\right)$, то гипотеза нормальности распределения случайных величин отклоняется.

В [277] предпринята попытка использовать вместо обычного коэффициента корреляции коэффициент ранговой корреляции Спирмена (см. раздел 5.2.2.2.2).

Задача 119. *Проверить гипотезу нормальности распределения вероятностей случайных величин в условиях задачи* 107 *критерием Лина–Мудхолкара при достоверности* $\alpha = 0{,}95$.

Имеем $\bar{x} = 4{,}8$; $s^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 = 24{,}4$.

Вычислим для примера значения $\bar{x}_1$ и s_1^2:

$$\bar{x}_1 = \frac{1}{10-1}\cdot[10\cdot 4{,}8 - (-1)] = 5{,}444;$$

$$s_1^2 = \frac{1}{10-2}\cdot\left[(10-1)\cdot 24{,}4 - \frac{10}{9}\cdot(-1-4{,}8)^2\right] = 22{,}778.$$

Далее по аналогии вычисляем пары значений $\bar{x}_i\left(s_i^2\right)^{\frac{1}{3}}$. Вычисления сводим в таблицу:

i	x_i	$\bar{x}_i$	s_i^2	$y_i = (s_i^2)^{\frac{1}{3}}$	$(\bar{x}_i - \bar{\bar{x}}\,(y_i - \bar{y}))$
1	−1	5,444	22,778	2,835	−0,03440
2	0	5,333	24,250	2,894	0,00286
3	1	5,222	25,440	2,941	0,02200
4	2	5,111	26,361	2,976	0,02690
5	3	5,000	27,000	3,000	0,02200
6	5	4,777	27,444	3,016	−0,00326
7	6	4,666	27,250	3,009	−0,01644
8	7	4,585	26,777	2,992	−0,02250
9	10	4,222	23,694	2,872	0,00960
10	15	3,666	13,000	2,351	0,61103
		48,026		28,886	0,62733

Вычисляем

$$\bar{\bar{x}} = \frac{1}{n}\cdot\sum_{i=1}^{n}\bar{x}_i = 4{,}8026; \quad \bar{y} = \frac{1}{n}\cdot\sum_{i=1}^{n}y_i = \frac{1}{n}\cdot\sum_{i=1}^{n}\left(s_i^2\right)^{\frac{1}{3}} = 2{,}8886;$$

$$\left[\sum_i (y_i - \bar{y})^2\right]^{\frac{1}{2}} = 0{,}596768; \quad \left[\sum_i (\bar{x}_i - \bar{\bar{x}})^2\right]^{\frac{1}{2}} = 1{,}64265;$$

$$r = \frac{0{,}62733}{0{,}596768\cdot 1{,}64265} = 0{,}63995; \quad z = \frac{1}{2}\cdot\ln\frac{1+0{,}63995}{1-0{,}63995} = 0{,}75809.$$

Далее находим

$$\sigma^2(z) = \frac{3}{10} - \frac{7{,}324}{100} + \frac{53{,}005}{1000} = 0{,}279765; \quad \alpha_4(z) = -\frac{11{,}7}{10} + \frac{55{,}06}{100} = -0{,}6194.$$

Имеем при $\alpha = 0{,}95$ (критерий двусторонний): $u_{\frac{1+\alpha}{2}} = u_{0{,}975} = 1{,}96$.

Тогда $z(0{,}95) = 0{,}52893\cdot\left\{1{,}96 + \dfrac{1}{24}\cdot\left(1{,}96^3 - 3\cdot 1{,}96\right)\cdot(-0{,}6194)\right\} = 1{,}01418$.

Так как $|z| = 0{,}75809 < z(\alpha) = 1{,}01418$, нулевая гипотеза нормальности не отклоняется.

3.2.2.14. Критерий нормальности Мартинеса–Иглевича

Критерий применяется против симметричных альтернатив, отличающихся от нормального распределения „хвостами" или эксцессом. Критерий основан на отношении двух оценок дисперсии — обычной и так называемой „робастной" (устойчивой), двухвесовой.

„Робастная", устойчивая к выбросам, оценка дисперсии имеет вид [278]

$$\tilde{s}^2 = \frac{n \sum\limits_{|z_i|<1} (x_i - \tilde{x})^2 \left(1 - z_i^2\right)^4}{\left\{\sum\limits_{|z_i|<1} \left(1 - z_i^2\right)\left(1 - 5z_i^2\right)\right\}^2}, \quad \text{где} \quad z_i = \begin{cases} \dfrac{x_i - \tilde{x}}{9\,\mathrm{med}|x_i - \tilde{x}|}, & \text{при } |z_i| < 1; \\ 0, & \text{при } |z_i| \geqslant 1. \end{cases}$$

$\tilde{x}$ — выборочная медиана, $\mathrm{med}(\ldots)$ — медиана ряда в скобках.

В качестве статистики используется отношение дисперсий [278]

$$I = \frac{s^2}{\tilde{s}^2} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)\tilde{s}^2} = \frac{1}{n(n-1)} \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 \left\{\sum\limits_{|z_i|<1} \left(1 - z_i^2\right)\left(1 - 5z_i^2\right)\right\}^2}{\sum\limits_{|z_i|<1} (x_i - \tilde{x})^2 (1 - z_i)^2}.$$

При справедливости нулевой гипотезы H_0 имеет место $M(I) = 0{,}982$. Критические значения статистики для уровней достоверности $\alpha = 0{,}90$ и $0{,}95$ могут быть найдены из соотношений

$$I^*(0{,}9) = 0{,}6376 - 1{,}1535n^* + 0{,}1266n^{*2};$$

$$I^*(0{,}95) = \begin{cases} 1{,}9065 - 2{,}5465n^* + 0{,}5652n^{*2}, & n > 50; \\ 0{,}7824 - 1{,}1021n^* + 0{,}1021n^{*2}, & n \leqslant 50; \end{cases}$$

где $n^* = \lg(n-1)$; $I^*(\alpha) = \lg[I(\alpha) - 0{,}982]$.

Этот критерий более мощен для альтернатив с длинными „хвостами", чем многие другие критерии.

Задача 120. *Проверить гипотезу нормальности распределения случайных величин в условиях задачи* 107 *критерием Мартинеса–Иглевича при доверительной вероятности* $\alpha = 0{,}95$.

Для ряда 1, 0, 1, 2, 3, 5, 6, 7, 10, 15 медианой является

$$\tilde{x} = \frac{x_5 + x_6}{2} = \frac{3+5}{2} = 4.$$

Составляем и ранжируем по величине ряд значений величин

$$|x_i - \tilde{x}|: \ 1, \ 1, \ 2, \ 2, \ 3, \ 3, \ 4, \ 5, \ 6, \ 11.$$

Вычисляем

$$A = \mathrm{med}|x_i - \tilde{x}| = \frac{3+3}{2} = 3.$$

Результаты расчетов сводим в таблицу (для упрощения записи приняты обозначения $\# = (1 - z_i^2 t) \cdot (1 - 5z_i^2)$ и $\#\# = (x_i - \tilde{x}) \cdot (1 - z_i^2)^4$):

i	z_i	z_i^2	$(1-z_i^2)^4$	#	$(x_i-\tilde{x})^2$	##
1	−5/27	0,0343	0,8697	0,8001	25	21,7420
2	−4/27	0,0220	0,9150	0,8707	16	14,6400
3	−3/27	0,0123	0,9515	0,9267	9	8,5630
4	−2/27	0,0055	0,9782	0,9672	4	3,9130
5	−1/27	0,0014	0,9945	0,9918	1	0,9945
6	1/27	0,0014	0,9945	0,9918	1	0,9945
7	2/27	0,0055	0,9782	0,9672	4	3,9130
8	3/27	0,0123	0,9515	0,9267	9	8,5630
9	6/27	0,0494	0,8353	0,7200	36	30,0710
10	11/27	0,1660	0,4838	0,1419	121	58,5400

Имеем

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 219{,}6; \qquad \sum_{|z_i|<1}(1-z_i^2)\cdot(1-5z_i^2) = 8{,}3041;$$

$$\sum_{|z_i|<1}(1-z_i^2)^4\cdot(x_i-\tilde{x})^2 = 151{,}934; \qquad I = \frac{1}{10\cdot 9}\cdot\frac{219{,}6\cdot 8{,}3041^2}{151{,}934} = 1.107;$$

$$I^*(0{,}95) = 1{,}9065 - 2{,}5465\cdot 0{,}95424 + 0{,}5652\cdot(0{,}95424)^2 = -8{,}8157\cdot 10^{-3};$$

$$I(0{,}95) = 10^{-8{,}8157\cdot 10^{-3}} + 0{,}982 = 1{,}962.$$

Так как $I = 1{,}107 < I(0{,}95) = 1{,}962$, нулевая гипотеза нормальности не отклоняется.

3.2.2.15. Критерий нормальности Д'Агостино

Д'Агостино [279, 280] предложил в качестве статистики для проверки нормальности распределения использовать отношение оценки Даутона для стандартного отклонения [281] (см. раздел 2.1.2.1.6.3) к выборочному стандартному отклонению, оцененному методом максимального правдоподобия,

$$D = \frac{T}{n^2 s},$$

где $$T = \sum_{i=1}^{n}\left\{i - \frac{n+1}{2}\right\}x_i, \quad x_1 \leqslant \ldots \leqslant x_n; \quad s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

Очевидно, что $\tilde{\sigma} = \dfrac{2\sqrt{\pi}\,T}{n(n-1)}$ является несмещенной оценкой стандартного отклонения σ.

Показано, что

$$M(D) = \frac{n-1}{2\sqrt{2\pi n}}\frac{\Gamma\left(\dfrac{n-1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)} \approx \frac{1}{2\sqrt{\pi}}\sqrt{\frac{n-1}{n}}\left[1 + \frac{1}{4(n-1)} + \frac{1}{32(n-1)^2} - \frac{5}{128(n-1)^3}\right] \approx$$
$$\approx \left(2\sqrt{\pi}\right)^{-1} = 0{,}28209479;$$

$$[D(D)]^{\frac{1}{2}} = \left(\frac{12\sqrt{3}-37+2\pi}{24n\pi}\right)^{\frac{1}{2}} = \frac{0{,}02998598}{\sqrt{n}};$$

$$\alpha_3(D) = -\frac{8{,}463}{\sqrt{n}} \text{ — коэффициент ассиметрии;}$$

$$\alpha_4(D) = \frac{107{,}9}{n} \text{ — коэффициент эксцесса.}$$

В качестве статистики критерия Д'Агостино используется величина

$$Y = \sqrt{n}\,\frac{D - 0{,}28209479}{0{,}02998598}.$$

критические значения которой приведены в [279, 280] и частично воспроизведены в табл. 77. Гипотеза нормальности принимается, если $Y_1(\alpha) \leqslant Y \leqslant Y_2(\alpha)$, где $Y_1(\alpha)$ и $Y_2(\alpha)$ — критические значения статистики Y при достоверности α.

Таблица 77

Критические значения критерия Д'Агостино [279, 280]

n	$\alpha = 0{,}90$		$\alpha = 0{,}95$		$\alpha = 0{,}99$	
	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2
10	−2,62	0,235	−3,25	0,299	−4,66	0,385
12	−2,58	0,329	−3,20	0,381	−4,63	0,479
14	−2,53	0,399	−3,16	0,460	−4,57	0,555
16	−2,50	0,459	−3,12	0,526	−4,52	0,613
18	−2,47	0,515	−3,08	0,574	−4,47	0,667
20	−2,44	0,565	−3,04	0,628	−4,41	0,720
22	−2,41	0,609	−3,01	0,677	−4,36	0,775
24	−2,39	0,648	−2,98	0,720	−4,32	0,822
26	−2,37	0,682	−2,96	0,760	−4,27	0,867
28	−2,35	0,714	−2,93	0,797	−4,23	0,910
30	−2,33	0,743	−2,91	0,830	−4,19	0,941
32	−2,32	0,770	−2,88	0,862	−4,16	0,983
34	−2,30	0,794	−2,86	0,891	−4,12	1,020
36	−2,29	0,816	−2,85	0,917	−4,09	1,050
38	−2,28	0,837	−2,83	0,941	−4,06	1,080
40	−2,26	0,857	−2,81	0,964	−4,03	1,110
42	−2,25	0,875	−2,80	0,986	−4,00	1,140
44	−2,24	0,892	−2,78	1,010	−3,98	1,170
46	−2,23	0,908	−2,77	1,020	−3,95	1,190
48	−2,22	0,923	−2,75	1,040	−3,93	1,220
50	−2,21	0,937	−2,74	1,060	−3,91	1,240
60	−2,17	0,997	−2,68	1,130	−3,81	1,340
70	−2,14	1,050	−2,64	1,190	−3,73	1,420
80	−2,11	1,080	−2,60	1,240	−3,67	1,480
90	−2,09	1,120	−2,57	1,280	−3,61	1,540
100	−2,07	1,140	−2,54	1,310	−3,57	1,590
150	−2,00	1,230	−2,45	1,420	−3,41	1,750
200	−1,96	1,290	−2,39	1,500	−3,30	1,850
250	−1,93	1,330	−2,35	1,540	−3,23	1,930
300	−1,91	1,360	−2,32	1,570	−3,17	1,980
350	−1,89	1,380	−2,29	1,610	−3,13	2,030
400	−1,87	1,400	−2,27	1,630	−3,09	2,060
450	−1,86	1,410	−2,25	1,650	−3,06	2,090
500	−1,85	1,420	−2,24	1,676	−3,04	2,110
600	−1,83	1,440	−2,21	1,690	−3,00	2,150
700	−1,82	1,460	−2,20	1,710	−2,97	2,180
800	−1,81	1,470	−2,18	1,730	−2,94	2,210
900	−1,80	1,480	−2,17	1,740	−2,92	2,310
1000	−1,79	1,490	−2,16	1,750	−2,91	2,350

Критерий Д'Агостино следует применять, когда нет сведений об альтернативном распределении. Он показывает хорошую мощность против большого спектра альтернатив, по мощности немного уступая критерию Шапиро–Уилка [282].

Задача 121. *В условиях задачи* 107 *проверить гипотезу нормальности распределения вероятностей случайных величин критерием Д'Агостино при достоверности* $\alpha = 0{,}95$ $(s = 4{,}686)$.

Вычисляем $T = \sum_{i=1}^{n} \left\{ i - \frac{n+1}{2} \right\} \cdot x_i = \left(1 - \frac{11}{2}\right) \cdot (-1) + \ldots + \left(9 - \frac{11}{2}\right) \cdot 15 = 129.$

Далее получаем

$$\tilde{\sigma} = \frac{2 \cdot \sqrt{\pi} \cdot 129}{10 \cdot 9} = 5{,}081; \quad D = \frac{129}{100 \cdot 4{,}686} = 0{,}275288;$$

$$Y = \sqrt{10} \cdot \frac{0{,}275288 - 0{,}28209479}{0{,}02998598} = -0{,}71783.$$

Из табл. 77 для $n = 10$ и $\alpha = 0{,}95$ находим $y_1(0{,}95) = -3{,}25$ и $y_2(0{,}95) = 0{,}299$. Так как $y_1(0{,}95) = -3{,}25 < Y = -0{,}71783 < y_2(0{,}95) = 0{,}299$, гипотеза нормальности распределения случайных величин не отклоняется.

3.2.2.16. Критерии асимметрии и эксцесса

Если распределение случайных величин нормально, то его коэффициент асимметрии $\alpha_3 = 0$ и коэффициент эксцесса $\alpha_4 = 3$. Напомним, что выборочные оценки коэффициентов асимметрии и эксцесса равны

$$\alpha_3 = \frac{1}{ns_3} \sum_{i=1}^{n} (x_i - \bar{x})^3; \quad \alpha_4 = \frac{1}{ns^4} \sum_{i=1}^{n} (x_i - \bar{x})^4; \quad s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Так как значения $\alpha_3 = 0$ и $\alpha_4 = 3$ могут иметь место и для распределений, отличных от нормального, то критерии этого раздела следует воспринимать как критерии установления отклонения от нормальности распределения (но не установления нормальности).

Известно [25], что $\mathbf{M}(\alpha_3) = 0$; $\mathbf{D}(\alpha_3) = \dfrac{6(n-2)}{(n+1)(n+3)} \approx \dfrac{6}{n}\left(1 - \dfrac{12}{2n+7}\right)$.

В [283] показано, что распределение α_3 достаточно быстро стремится к нормальному. Для α_4 справедливы соотношения

$$\mathbf{M}(\alpha_4) = 3 - \frac{6}{n+1}; \quad \mathbf{D}(\alpha_4) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} = \frac{24}{n}\left(1 - \frac{225}{15n+224}\right).$$

Распределение α_4 медленно стремится к нормальному [283].

Рассмотрим использование критерия α_3 для установления отклонения эмпирического распределения от нормального. Таблицы процентных точек распределения α_3 приведены в [25]. При $n > 200$ может быть рекомендован грубый критерий: если

$$\alpha_3 = \frac{1}{ns^3} \sum_{i=1}^{n} (x_i - \tilde{x})^3 \geqslant \frac{6}{n},$$

то нормальность распределения отклоняется.

На практике применяются нормализующие преобразования для α_3. Рассмотрим некоторые из них. В [284] предложена аппроксимация $\xi = \delta \ln\left(x + \sqrt{x^2+1}\right)$, где $x = \dfrac{\alpha_3}{\lambda}$, которая при $n \to \infty$ распределена как стандартная нормальная величина (δ, λ — коэффициенты, приведенные в табл. 78).

Таблица 78

Значения коэффициентов δ и $\frac{1}{\lambda}$ [284]

n	δ	$1/\lambda$	n	δ	$1/\lambda$	n	δ	$1/\lambda$
8	5,563	0,3030	44	3,117	0,9795	125	4,272	1,1250
9	4,260	0,4080	45	3,131	0,9840	130	4,336	1,1285
10	3,734	0,4794	46	3,160	0,9882	135	4,398	1,1318
11	3,447	0,5839	47	3,161	0,9923	140	4,460	1,1348
12	3,270	0,5781	48	3,176	0,9963	145	4,521	1,1377
13	3,151	0,6153	49	3,192	1,0001	150	4,582	1,1403
14	3,069	0,6743	50	3,207	1,0038	155	4,641	1,1428
15	3,010	0,6753	52	3,237	1,0108	160	4,700	1,1452
16	2,968	0,7001	54	3,268	1,0174	165	4,758	1,1474
17	2,937	0,7224	56	3,298	1,0235	170	4,816	1,1496
18	2,915	0,7426	58	3,329	1,0293	175	4,873	1,1516
19	2,900	0,7610	60	3,359	1,0348	180	4,929	1,1535
20	2,890	0,7779	62	3,389	1,0400	185	4,985	1,1553
21	2,870	0,7940	64	3,420	1,0449	190	5,040	1,1570
22	2,882	0,8078	66	3,450	1,0459	195	5,094	1,1586
23	2,882	0,8211	68	3,480	1,0540	200	5,148	1,1602
24	2,884	0,8336	70	3,510	1,0581	210	5,255	1,1631
25	2,889	0,8452	72	3,540	1,0621	220	5,359	1,1657
26	2,895	0,8561	74	3,569	1,0659	230	5,461	1,1681
27	2,902	0,8664	76	3,599	1,0695	240	5,561	1,1704
28	2,910	0,8760	78	3,628	1,0730	250	5,660	1,1724
29	2,920	0,8851	80	3,657	1,0763	260	5,757	1,1744
30	2,930	0,8938	82	3,686	1,0795	270	5,853	1,1761
31	2,941	0,9020	84	3,715	1,0825	280	5,946	1,1779
32	2,952	0,9097	86	3,744	1,0854	290	6,039	1,1793
33	2,964	0,9171	88	3,772	1,0882	300	6,130	1,1808
34	2,977	0,9241	90	3,801	1,0909	350	6,567	1,1868
35	2,990	0,9308	92	3,829	1,0934	400	6,976	1,1914
36	3,003	0,9372	94	3,857	1,0959	450	7,363	1,1950
37	3,016	0,9433	96	3,885	1,0983	500	7,731	1,1979
38	3,030	0,9492	98	3,913	1,1006	600	8,149	1,2023
39	3,044	0,9548	100	3,940	1,1028	700	9,054	1,2058
40	3,058	0,9601	105	4,009	1,1080	800	9,649	1,2078
41	3,073	0,9653	110	4,076	1,1128	900	10,271	1,2096
42	3,087	0,9702	115	4,142	1,1172	1000	10,738	1,2111
43	3,102	0,9750	120	4,207	1,1212			

В [285] предложена следующая нормализующая аппроксимация. Если

$$y = \alpha_3 \sqrt{\frac{(n+1)(n+3)}{6(n-2)}}; \quad \beta = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)};$$

$$\omega^2 = -1 + \sqrt{2(\beta-1)}; \quad \delta = \sqrt{\ln 10}; \quad \alpha = \left(\frac{2}{\omega^2-1}\right)^{\frac{1}{2}},$$

то величина $z = \delta \ln\left\{\frac{y}{\alpha} + \sqrt{\left(\frac{y}{\alpha}\right)^2 + 1}\right\}$ уже при $n > 25$ может быть аппроксимирована стандартным нормальным распределением.

Рассмотрим теперь преобразование для коэффициента эксцесса α_4 [286]. Распределение α_4 может быть аппроксимировано распределением χ^2 с f степенями свободы при

$$f = 6 + \frac{8}{c}\left\{\frac{2}{c} + \left(1 + \frac{4}{c^2}\right)^{\frac{1}{2}}\right\}, \quad \text{где} \quad c = \frac{6(n^2 - 5n + 2)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}.$$

В [286] предложено весьма эффективное нормализующее преобразование для коэффициента эксцесса α_4. Алгоритм его построения заключается в следующем. Если $x = \dfrac{\alpha_4 - \mathbf{M}(\alpha_4)}{\sqrt{\mathbf{D}(\alpha_4)}}$, то случайная величина

$$d = \left[1 - \frac{2}{9f} - \left\{\frac{1 - \frac{2}{f}}{1 + x\sqrt{\frac{2}{f-4}}}\right\}^{\frac{1}{3}}\right]\left(\frac{2}{9f}\right)^{-\frac{1}{2}}$$

аппроксимируется стандартным нормальным распределением $N(0,1)$ уже при $n > 20$.

Нормализующие преобразования позволяют использовать таблицы (или аппроксимации) стандартного нормального распределения для проверки отклонения от нормальности.

Мощность критерия проверки отклонения от нормальности может быть повышена применением так называемого комбинированного K^2-критерия [284, 289]

$$K^2 = X^2(\alpha_3) + X^2(\alpha_4),$$

где $X(\alpha_3)$ и $X(\alpha_4)$ — стандартные нормальные эквиваленты распределений α_3 и α_4.

Статистика K^2 имеет χ^2-распределение с $f = 2$ степенями свободы. Другая форма комбинированного критерия исследована в [287, 288]. Если $q_1 = 2p\ (\alpha_3 < a_3)$ и $q_2 = 2p\ (\alpha_4 < a_4)$ (a_3 и a_4 — выборочные оценки параметров α_3 и α_4 соответственно), то статистика $q = -2\ln q_1 q_2$ имеет χ^2-распределение с $f = 4$.

Задача 122. *В условиях задачи* 107 *проверить гипотезу нормальности распределения случайных величин критериями асимметрии и эксцесса на уровне достоверности* $\alpha = 0{,}90$.

Отметим, что задача носит демонстрационный характер, так как применение критериев такого типа требует объемов выборки не менее 50.

Имеем

$$\bar{x} = 4{,}8; \quad s^2 = 21{,}96 \ (s = 4{,}686); \quad \sum_{i=1}^{n}(x_i - \bar{x})^3 = 8{,}25{,}84;$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^4 = 13523{,}952; \quad \alpha_3 = \frac{1}{10 \cdot 4{,}686^3} \cdot 825{,}84 = 0{,}80258;$$

$$\alpha_4 = \frac{1}{10 \cdot 21{,}96^2} \cdot 13523{,}452 = 2{,}804393.$$

Вычисляем далее

$$\mathbf{D}(\alpha_3) = \frac{6 \cdot 8}{11 \cdot 13} = 0{,}33566 \quad (\sqrt{\mathbf{D}(\alpha_3)} = 0{,}57936);$$

$$\mathbf{M}(\alpha_4) = 3 - \frac{6}{11} = 2{,}4545; \quad \mathbf{D}(\alpha_4) = \frac{24 \cdot 10 \cdot 8 \cdot 7}{11^2 \cdot 13 \cdot 15} = 0{,}56961 \quad (\sqrt{\mathbf{D}(\alpha_4)} = 0{,}75473).$$

Грубый критерий $a_3 = 0{,}8025 > \sqrt{D(\alpha_3)} = 0{,}579$ отклоняет гипотезу нормальности (это применимо для $n > 200$, поэтому в нашем случае применение критерия некорректно; однако наша цель — демонстрация техники вычисления — достигается).

Вычислим нормализующие преобразования. Рассмотрим сначала ξ-преобразование. Из табл. 78 для $n = 10$ находим $\delta = 3{,}734$ и $\dfrac{1}{\lambda} = 0{,}4794$.

Далее имеем

$$x = \alpha_3 \cdot 0{,}4794 = 0{,}80258 \cdot 0{,}4794 = 0{,}384757; \quad \xi = 3{,}734 \cdot \ln(0{,}384757 + 1{,}071465) = 1{,}4034.$$

Рассмотрим теперь z-преобразование:

$$y = 0{,}80258 \cdot \sqrt{\frac{11 \cdot 13}{6 \cdot 8}} = 1{,}38527416; \quad \beta = \frac{3 \cdot (100 + 270 - 70)}{8 \cdot 15 \cdot 17 \cdot 19} \cdot 3{,}3204334;$$

$$\omega^2 = -1 + \sqrt{2 \cdot 3{,}3204334} = 1{,}5769879; \quad \delta = \left(\ln\sqrt{1{,}5769879}\right)^{-\frac{1}{2}} = 2{,}0953804;$$

$$\alpha = \left(\frac{2}{0{,}5769879}\right)^{\frac{1}{2}} = 1{,}8617941;$$

$$z = 2{,}0953804 \cdot \ln\left\{\frac{1{,}38527416}{1{,}8617941} + \sqrt{\left(\frac{1{,}38527416}{1{,}8617941}\right)^2 + 1}\right\} = 1{,}4424244.$$

Теперь рассмотрим преобразование для α_4. Сначала вычислим χ^2-преобразование. Имеем

$$c = \frac{6 \cdot (100 - 50 + 2)}{17 \cdot 19} \cdot \sqrt{\frac{6 \cdot 13 \cdot 15}{10 \cdot 8 \cdot 7}} = 1{,}3962;$$

$$f = 6 + \frac{8}{1{,}3962} \cdot \left[\frac{2}{1{,}3962} + \left(1 + \frac{4}{1{,}3962}\right)^{\frac{1}{2}}\right] = 25{,}47.$$

Следовательно, величина α_4 может быть аппроксимирована χ^2-распределением с $f \approx 25{,}5$ степенями свободы.

Наконец, рассмотрим нормализующее преобразование для α_4. Имеем

$$\mathbf{M}(\alpha_4) = 2{,}4545; \quad \mathbf{D}(\alpha_4) = 0{,}569161 \ (\sqrt{\mathbf{D}(\alpha_4)} = 0{,}75473); \quad x = \frac{2{,}804393 - 2{,}4545}{0{,}75473} = 0{,}4636;$$

$$d = \left(\frac{2}{9 \cdot 25{,}47}\right)^{-\frac{1}{2}} \cdot \left[1 - \frac{2}{9 \cdot 25{,}47} - \left\{\frac{1 - \dfrac{2}{25{,}47}}{1 + 0{,}4636 \cdot \sqrt{\dfrac{2}{21{,}47}}}\right\}^{\frac{1}{3}}\right] = 0{,}644.$$

Используем нормализующие преобразования для α_3 и α_4 для оценки отклонения эмпирического распределения отнормального ($u_{0,95} = 1{,}645$ — 95%-я квантиль стандартного нормального распределения. Так как критерий двустронний, при $\alpha = 0{,}90$ следует применять $u_{\frac{1+\alpha}{2}} = u_{\frac{1+0{,}90}{2}} = u_{0{,}95}$).

Имеем $z = 1{,}442 < u_{0{,}95} = 1{,}645$. Следовательно, гипотеза нормальности по коэффициенту асимметрии не отклоняется.

По коэффициенту эксцесса имеем $d = 0{,}664 < u_{0{,}95} = 1{,}645$ и гипотеза нормальности также не отклоняется.

Рассмотрим теперь комбинированный критерий. Нормализующие преобразования дают: $X^2(\alpha_3) = (1{,}403)^2 = 1{,}9684$; $X^2(\alpha_4) = (0{,}6498)^2 = 0{,}4222$ и $K^2 = X^2(\alpha_3) + X^2(\alpha_4) = 2{,}3906$. Из табл. 55 для $f = 2$ имеем критическое значение $\chi^2_{0,9}(2) = 4{,}605$. Так как $K^2 = 2{,}39 < \chi^2_{0,9}(f = 2) = 4{,}605$, гипотеза нормальности распределения случайных величин не отклоняется.

3.2.2.17. Критерий характеристической функции (критерий Муроты–Такеучи)

Если x_i ($i = 1, \ldots, n$) — выборка из распределения с характеристической функцией $c(t)$, то эмпирическая характеристическая функция определяется как

$$c_n(t) = \frac{1}{n}\sum_{i=1}^{n}\exp(itx_i).$$

В [290–292] исследована возможность применения характеристической функции для проверки различных гипотез согласия. Для проверки сложной гипотезы нормальности (когда параметры распределения не известны заранее) Мурота и Такеучи [293] предложили использовать „стьюдентизированную“ форму

$$\tilde{c}_n(t) = c_n\left(\frac{t}{s}\right), \quad \text{где} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Введем обозначения $a(t) = |c(t)|^2$ и $a_n(t) = |c_n(t)|^2$. В общем случае модуль характеристической функции $|c(t)|$ не определяет форму распределения, но для нормального распределения есть исключение. Распределение с характеристической функцией $c(t)$ нормально тогда и только тогда, когда величина $\sqrt{-\log|c(t)|^2}$ линейна при любом $t \geqslant 0$ [293].

Для оценки „стьюдентизированного“ модуля характеристической функции используется соотношение [293]

$$\tilde{a}_n(t) = \frac{1}{n} + \frac{2}{n^2}\sum_{j<k}\cos\left[\frac{t(x_j - x_k)}{s}\right].$$

Распределения величин $a_n(t)$ и $\tilde{a}_n(t)$ медленно нормализуются с ростом n. Лучшее приближение к нормальному распределению реализуется при $t = 1 \div 1{,}5$.

Гипотеза нормальности не отклоняется, если $a_1 \leqslant \tilde{a}_n(t) \leqslant a_2$, где a_1 и a_2 — граничные значения, зависящие от объема выборки и принятого уровня значимости (приведены в табл. 79).

Таблица 79

Критические значения a_1 и a_2 критерия Муроты–Такеучи на уровне значимости $\alpha = 0{,}10$ [293]

n	$t = 0{,}5$		$t = 1{,}0$		$t = 1{,}5$		$t = 2{,}0$	
	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
10	0,7934	0,8022	0,3604	0,4527	0,0540	0,2792	0,0027	0,2684
15	0,7871	0,7966	0,3512	0,4440	0,0545	0,2544	0,0022	0,1968
20	0,7841	0,7935	0,3475	0,4365	0,0555	0,2349	0,0022	0,1640
35	0,7805	0,7888	0,3462	0,4192	0,0619	0,1945	0,0025	0,1072
50	0,7793	0,7866	0,3466	0,4094	0,0660	0,1760	0,0024	0,0842

Против асимметричных альтернатив критерий лучше, чем критерии асимметрии и эксцесса (см. раздел 3.2.2.16). Критерий демонстрирует результаты не хуже, чем критерий Шапиро–Уилка (см. раздел 3.2.2.1), но он может быть использован против более широкого круга альтернатив.

Задача 123. *Проверить гипотезу нормальности распределения случайных величин для данных задачи* 107 *критерием Муроты–Такеучи на уровне значимости* $\alpha = 0{,}1$.

Имеем $s^2 = 24{,}4$ ($s = 4{,}3996$); $\bar{x} = 4{,}8$.

Будем использовать значение $t = 1$. Результаты расчетов сведены в таблицу:

j	k	$\frac{x_j - x_k}{s}$	$\cos\left(\frac{x_j - x_k}{s}\right)$	j	k	$\frac{x_j - x_k}{s}$	$\cos\left(\frac{x_j - x_k}{s}\right)$
1	2	−0,20244	0,9796	3	10	−2,8342	−0,9531
1	3	−0,4049	0,9191	4	5	−0,2024	0,9796
1	4	−0,6073	0,8212	4	6	−0,6073	0,8212
1	5	−0,8098	0,6896	4	7	−0,8098	0,6897
1	6	−1,2146	0,3486	4	8	−1,1022	0,5300
1	7	−1,4171	0,0308	4	9	−1,6195	−0,0487
1	8	−1,6195	−0,0487	4	10	−2,6318	−0,8728
1	9	−2,2269	−0,6100	5	6	−0,4049	0,9191
1	10	−3,2391	−0,9952	5	7	−0,6073	0,8212
2	3	−0,2024	0,9796	5	8	−0,8098	0,6897
2	4	−0,4049	0,9191	5	9	−1,4171	0,1531
2	5	−0,6073	0,8212	5	10	−2,4293	−0,7569
2	6	−1,0122	0,5300	6	7	−0,2024	0,9796
2	7	−1,2147	0,3486	6	8	−0,4049	0,9191
2	8	−1,4171	0,1531	6	9	−1,0122	0,5300
2	9	−2,0244	−0,4382	6	10	−2,0244	−0,4382
2	10	−3,0366	−0,9945	7	8	−0,2024	0,9796
3	4	−0,2024	0,9796	7	9	−0,8098	0,6896
3	5	−0,4049	0,9191	7	10	−1,8220	−0,2486
3	6	−0,8098	0,6897	8	9	−0,6073	0,8212
3	7	−1,0122	0,5300	8	10	−1,6195	−0,0487
3	8	−1,2146	0,3486	9	10	−1,0122	0,5300
3	9	−1,8220	−0,2485				
		$\sum$	7,6724				7,6857

Суммируя, находим $\sum\limits_{j<k} \cos\left(\frac{x_j - x_k}{s}\right) = 15{,}3581.$

Окончательно получаем

$$\tilde{a}_n(1) = \frac{1}{n} + \frac{2}{n^2} \cdot \sum_{j<k} \cos\left(\frac{x_j - x_k}{s}\right) = \frac{1}{10} + \frac{2}{100} \cdot 15{,}3581 = 0{,}40716.$$

Из табл. 79 для $t = 1$ и $n = 10$ имеем $a_1 = 0{,}3604$ и $a_2 = 0{,}4527$.

Так как $a_1 = 0{,}3604 < \tilde{a}_n(1) = 0{,}40716 < a_2 = 0{,}4527$, гипотеза нормальности распределения вероятностей случайных величин не отклоняется.

3.2.2.18. Критерии проверки нормальности распределения по совокупности независимых выборок малого объема

Исследователь или испытатель чаще всего встречается на практике с ситуацией, когда в его распоряжении находится совокупность малых выборок, поступающих последовательно в течение длительного промежутка времени наблюдений или испытаний, например, результаты испытаний малых партий изделий по мере их изготовления. В этом случае появляется необходимость проверки нормальности распределения случайной величины по совокупности выборок малого объема, в которых средние и дисперсии в общем случае могут быть разными. Различные методы установления нормальности распределения наблюдаемой случайной величины в такой ситуации рассматриваются в настоящем разделе.

3.2.2.18.1. Применение критерия Шапиро–Уилка [1])

Пусть имеется k независимых выборок объема n_i каждая ($i = 1, \ldots, k$) и $x_{i1} \leqslant x_{i2} \leqslant \ldots \leqslant x_{in_i}$ для каждой i-й выборки. Для каждой выборки вычислим статистику

$$W_i = \frac{\left\{ \sum\limits_{j=1}^{\left[\frac{n}{2}\right]} a_j (x_{i,n_i-j+1} - x_{ij}) \right\}^2}{\sum\limits_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}.$$

Здесь $[\ldots]$ — наибольшее целое число.

Пусть $\alpha_i = F(W_i)$ — уровень значимости W_i-статистики; u_{α_i} — α-квантиль стандартного нормального распределения; $U = \frac{1}{\sqrt{k}} \sum\limits_{i=1}^{k} u_{\alpha_i}$; $c_i = -2 \ln \alpha_1$ — верхняя α-точка распределения χ^2 с $f = 2$ степенями свободы и $c = \sum\limits_{i=1}^{k} c_i$.

Идея проверки нормальности распределения заключается в том, что все u_{α_i} и U имеют стандартное нормальное распределение, а c имеет распределение χ^2 с $f = 2k$ степенями свободы [294].

Ранее [242] была найдена аппроксимация

$$u_{\alpha_i} = \gamma + \eta \ln \frac{W_i - \varepsilon}{1 - W_i},$$

коэффициенты которой приведены в табл. 65.

Статистика $U = \frac{1}{\sqrt{k}} \sum\limits_{i=1}^{k} u_{\alpha i}$ имеет стандартное нормальное распределение. Отклонение распределения от нормального приводит к уменьшению значений W_i. Сдвиг значений u_{α_i} к малым отрицательным величинам сигнализирует об этом. Поэтому статистика критерия должна быть интерпретирована в терминах одностороннего критерия, т. е. если $U < u_\alpha$, то на уровне значимости α гипотеза нормальности должна быть отклонена.

Во втором варианте критерия вычисляются значения $c_i = -2 \ln \alpha_i$ и $c = \sum\limits_{i=1}^{k} c_i$. Если $c > \chi^2_\alpha(2k)$, то гипотеза нормальности распределения не отклоняется ($c > \chi^2_\alpha(2k)$, критическое значение χ^2-статистики на уровне значимости α с f степенями свободы).

3.2.2.18.2. Применение критерия Саркади [2])

В [295] предложено применение модифицированного критерия Саркади для проверки нормальности распределения случайных величин по совокупности выборок малого объема. Алгоритм вычисления статистики критерия включает в себя

[1]) См. раздел 3.2.2.1 — основной критерий.
[2]) См. раздел 3.2.2.12 — основной критерий.

вычисление для каждой i-й выборки величин

$$x_{ij} - \frac{1}{n_i + \sqrt{n_i}} \sum_{k=1}^{n_i} x_{ik} - \frac{1}{\sqrt{n_i}+1} x_{im}, \quad j = 1, 2, \ldots, m-1;$$

$$x_{i,j+1} - \frac{1}{n_i + \sqrt{n_i}} \sum_{k=1}^{n_i} x_{ik} - \frac{1}{\sqrt{n_i}+1} x_{im}, \quad j = m, \ldots, n_i - 1; \quad i = 1, \ldots, k$$

и

$$z_j^i = \left\{ \frac{1}{n_i - j - 1} \left(y_{j+1}^i + \ldots + y_{n_i-1}^i \right) \right\}^{-\frac{1}{2}} y_j^i, \quad j = 1, \ldots, n_i - 2; \quad i = 1, \ldots, k.$$

Величины z_j^i имеют t-распределение Стьюдента с функцией распределения вероятностей $F_{n_i-j-1}(z)$.

Находим величины $r_j^i = F_{n_i-j-1}\left(z_j^i\right)$ и затем объединяем все величины r_j^i в одну выборку объема $\sum(n_i - 2)$. К ней применяется „гладкий" критерий Неймана для проверки равномерности распределения величин r_{ji} на отрезке $[0, 1]$ со статистикой

$$P_4^2 = \frac{1}{\sum\limits_{i=1}^{k}(n_i - 2)} \sum_{i=1}^{4} \sum_{j=1}^{\sum\limits_{i=1}^{k}(n_i-2)} \left[\varphi_i(r_j)\right]^2,$$

где

$$\varphi_1(y) = \sqrt{12}\left(y - \frac{1}{2}\right); \quad \varphi_2(y) = \sqrt{5}\left[6\left(y - \frac{1}{2}\right)^2 - \frac{1}{2}\right];$$

$$\varphi_3(y) = \sqrt{7} \cdot \left[20\left(y - \frac{1}{2}\right)^3 - 3\left(y - \frac{1}{2}\right)\right];$$

$$\varphi_4(y) = 210\left(y - \frac{1}{2}\right)^4 - 45\left(y - \frac{1}{2}\right)^2 + \frac{9}{8}.$$

Если $P_4^2 < \chi_\alpha^2(4)$, то гипотеза нормальности распределения не отклоняется.

3.2.2.18.3. Критерий Смирнова [12]

Критерий основан на статистике

$$\tau = \frac{x_{ij} - \bar{x}_i}{s_i},$$

где x_{ij} — j-е наблюдение i-й выборки; $\bar{x}_i, s_i$ — соответственно среднее значение и стандартное отклонение i-й выборки.

Распределение статистики табулировано в [297]. При $n_i = n = 4$ τ имеет равномерное распределение, а при $n \neq 4$ статистика $t = \dfrac{\tau - \sqrt{n-2}}{\sqrt{n - 1 - \tau^2}}$ имеет t-распределение Стьюдента с $f = n - 2$ степенями свободы. Проверкой соответствия распределения величины t распределению Стьюдента и устанавливается нормальность распределения случайной величины в исходных выборках. Идея алгоритма такая же, как и в предыдущем критерии Саркади — проверяется равномерность распределения величин τ $(n = 4)$ или согласие с распределением Стьюдента величин t $(n \neq 4)$.

Задача 124. *Имеются 5 выборок по 4 наблюдения в каждой.*

Необходимо проверить гипотезу нормальности распределения случайных величин по приведенной совокупности выборок при достоверности $\alpha = 0{,}95$.

$$\begin{array}{lllll} x_{11} = -1, & x_{21} = 6, & x_{31} = 7, & x_{41} = 0, & x_{51} = -3, \\ x_{12} = 0, & x_{22} = 9, & x_{32} = 11, & x_{42} = 1, & x_{52} = -1, \\ x_{13} = 1, & x_{23} = 11, & x_{33} = 15, & x_{43} = 4, & x_{53} = 0, \\ x_{14} = 2, & x_{24} = 13, & x_{34} = 16, & x_{44} = 5, & x_{54} = 1. \end{array}$$

Критерий Шапиро–Уилка (3.2.2.18.1)

Для 1-й выборки находим $\bar{x}_1 = 0{,}5$; $\sum\limits_{j=1}^{4}(x_{1j} - \bar{x}_1) = 5$. Из табл. 63 для $n = 4$ находим $a_1 = 0{,}6872$ и $a_2 = 0{,}1677$. Тогда имеем

$$W_1 = \frac{\{0{,}6872 \cdot (x_{14} - x_{11}) + 0{,}1677 \cdot (x_{13} - x_{12})\}^2}{\sum\limits_{j=1}^{4}(x_{1j} - \bar{x}_1)^2} = 0{,}9939.$$

Из табл. 65 для $n = 4$ имеем $\gamma = -1{,}107$, $\eta = 0{,}714$ и $\varepsilon = 0{,}6297$.

Далее вычисляем

$$u_{\alpha_1} = -1{,}107 + 0{,}714 \cdot \ln \frac{0{,}9939 - 0{,}6297}{1 - 0{,}9939} = 1{,}8128.$$

Из табл. 1 находим α_1, соответствующее квантили стандартного нормального распределения $u_{\alpha_1} = 1{,}8128$ — это $\alpha_1 = 0{,}965$.

Далее по аналогии находим

$$\begin{array}{llll} \bar{x}_2 = 9{,}75; & W_2 = 0{,}9899; & u_{\alpha_2} = 1{,}445; & \alpha_2 = 0{,}926; \\ \bar{x}_3 = 12{,}25; & W_3 = 0{,}9261; & u_{\alpha_3} = -0{,}1152; & \alpha_3 = 0{,}460; \\ \bar{x}_4 = 2{,}5; & W_4 = 0{,}9127; & u_{\alpha_4} = -0{,}267; & \alpha_4 = 0{,}394; \\ \bar{x}_5 = -0{,}75; & W_5 = 0{,}9721; & u_{\alpha_5} = 0{,}6832; & \alpha_5 = 0{,}751 \end{array}$$

и

$$U = \frac{1}{\sqrt{5}} \cdot \sum_{i=1}^{5} u_{\alpha_i} = 1{,}5915.$$

Из табл. 1 для $\alpha = 0{,}95$ находим $u_{0{,}95} = 1{,}645$.

Так как $U = 1{,}5915 < u_{0{,}95} = 1{,}645$, гипотеза нормальности распределения случайных величин не отклоняется.

Далее вычисляем

$$c = \sum_{i=1}^{5}(-2\ln\alpha_i) = -2 \cdot (\ln 0{,}965 + \ldots + \ln 0{,}751) = 4{,}213.$$

Из табл. 55 находим $\chi^2_{0{,}95}(f = 10) = 18{,}307$.

Так как $c = 4{,}213 < \chi^2_{0{,}95}(10) = 18{,}307$, гипотеза нормальности не отклоняется.

Критерий Саркади (3.2.2.18.2)

Выбираем $m = n = 4$ и для первой выборки имеем

$$y'_1 = x_{11} - \frac{1}{4 + \sqrt{4}} \cdot \sum_{k=1}^{4} x_{ik} - \frac{1}{\sqrt{4} + 1} \cdot x_{14} = -1 - \frac{1}{6} \cdot 2 - \frac{1}{3} \cdot 2 = -2;$$

$$y'_2 = -1; \quad y'_3 = 0.$$

Вычисляем

$$z_1' = \left\{\frac{1}{4-1-1}\cdot\left({y_2'}^2+{y_3'}^2\right)\right\}^{-\frac{1}{2}}\cdot y_1' = \frac{-2}{\sqrt{\frac{1}{2}\cdot(1+0)}} = -2{,}828; \quad z_2' = -\infty.$$

Воспользовавшись функцией распределения Стьюдента, имеем (см. табл. 3.1а из [25]) $r_1' = F_2(-2{,}828) = 0{,}053$ и $r_2' = 0$. Далее по аналогии имеем

$$y_1^2 = x_{21} - \frac{1}{5}\cdot\sum_{k=1}^{4} x_{2k} - \frac{1}{3}\cdot x_{24} = -4{,}833; \quad y_2^2 = -1{,}833; \quad y_3^2 = 0{,}1667;$$

$$z_1^2 = -3{,}7131; \quad z_2^2 = -11; \quad r_1^2 = 0{,}032; \quad r_2^2 \approx 0;$$

$$y_1^3 = -6{,}5; \quad y_2^4 = -2{,}333; \quad y_3^4 = 0{,}6667; \quad z_1^4 = -3{,}5; \quad r_1^4 = 0{,}094; \quad r_2^4 = 0{,}089;$$

$$y_1^5 = -3{,}8333; \quad y_2^5 = -0{,}8333; \quad z_1^5 = -2{,}6919; \quad z_2^5 = -2{,}20; \quad r_1^5 = 0{,}06; \quad r_2^5 = 0{,}136.$$

Получаем ряд величин:

$$r_1^1 = 0{,}053; \quad r_2^1 = 0; \quad r_1^2 = 0{,}032; \quad r_2^2 = 0{,}0444; \quad r_2^3 = 0{,}175;$$

$$r_1^4 = 0{,}084; \quad r_2^4 = 0{,}089; \quad r_1^5 = 0{,}060; \quad r_2^5 = 0{,}136.$$

Находим

$$\varphi_1(r_1^1) = -1{,}548; \quad \varphi_1(r_2^1) = -1{,}732; \quad \varphi_1(r_1^2) = -1{,}621; \quad \varphi_1(r_2^2) = -1{,}580;$$

$$\varphi_1(r_2^3) = -1{,}126; \quad \varphi_1(r_1^4) = -1{,}406; \quad \varphi_1(r_2^4) = -1{,}424; \quad \varphi_1(r_1^5) = -1{,}524;$$

$$\varphi_1(r_2^5) = -1{,}261 \quad \text{и} \quad \sum \varphi_i^2 = \varphi_2(r_1^1) = 1{,}563;$$

$$\varphi_2(r_2^1) = 2{,}236; \quad \varphi_2(r_1^2) = 1{,}820; \quad \varphi_2(r_2^2) = 2{,}236; \quad \varphi_2(r_1^3) = 1{,}672;$$

$$\varphi_2(r_2^3) = 0{,}299; \quad \varphi_2(r_1^4) = 1{,}093; \quad \varphi_2(r_2^4) = 1{,}148; \quad \varphi_2(r_1^5) = 1{,}479;$$

$$\varphi_2(r_2^5) = 0{,}659 \quad \text{и} \quad \sum \varphi_2^2 = 23{,}774;$$

$$\varphi_3(r_1^1) = -1{,}178; \quad \varphi_3(r_2^1) = -2{,}645; \quad \varphi_3(r_1^2) = -1{,}709; \quad \varphi_3(r_2^2) = -2{,}645;$$

$$\varphi_3(r_1^3) = -1{,}398; \quad \varphi_3(r_2^3) = 0{,}763; \quad \varphi_3(r_1^4) = -0{,}319; \quad \varphi_3(r_2^4) = -0{,}411;$$

$$\varphi_3(r_1^5) = -1{,}015; \quad \varphi_3(r_2^5) = 0{,}337 \quad \text{и} \quad \sum \varphi_3^2 = 22{,}251.$$

Окончательно получаем

$$P_4^2 = \frac{1}{10}\cdot\sum_{i=1}^{4}\varphi_i^2 = 8{,}527.$$

Из табл. 55 имеем $\chi^2_{0,95}(4) = 9{,}488$.

Так как $P_4^2 = 8{,}527 < \chi^2_{0,95}(4) = 9{,}488$, гипотеза нормальности распределения случайных величин не отклоняется.

3.2.2.19. Сравнительная мощность различных критериев нормальности

В этом разделе представлены результаты исследования сравнительной мощности критериев нормальности распределения вероятностей случайных величин [243] для различных альтернативных распределений. В табл. 80 представлено ранжирование 21 критерия нормальности. Критерии по каждой альтернативе представлены в порядке предпочтения — от наибольшего (1) до наименьшего (21). В последней графе приведено общее ранжирование, соответствующее набранной сумме рангов. Табл. 80 может быть полезной ориентировкой для пользователя при выборе критерия проверки нормальности распределения вероятностей изучаемой случайной величины.

Таблица 80

Сравнение критериев проверки нормальности распределения случайных величин

Наименование критерия (раздел)	Характер альтернативного распределения					Ранг
	асимметричное		симметричное		≈ нормальное	
	$\alpha_4 < 3$	$\alpha_4 > 3$	$\alpha_4 < 3$	$\alpha_4 > 3$	$\alpha_4 \approx 3$	
Критерий Шапиро–Уилка (3.2.2.1)	1	1	3	2	2	1
Критерий K^2 (3.2.2.16)	7	8	10	6	4	2
Критерий Дарбина (3.1.2.7)	11	7	7	15	1	3
Критерий Д'Агостино (3.2.2.14)	12	9	4	5	12	4
Критерий α_4 (3.2.2.16)	14	5	2	4	18	5
Критерий Васичека (3.2.2.2)	2	14	8	10	10	6
Критерий Дэвида–Хартли–Пирсона (3.2.2.10)	21	2	1	9	1	7
Критерий χ^2 (3.1.1.1)	9	20	9	8	3	8
Критерий Андерсона–Дарлинга (3.1.2.4)	18	3	5	18	7	9
Критерий Филлибена (3.2.2.5)	3	12	18	1	9	10
Критерий Колмогорова–Смирнова (3.1.2.1)	16	10	6	16	5	11
Критерий Мартинеса–Иглевича (3.2.2.14)	10	16	13	3	15	12
Критерий Лина–Мудхолкара (3.2.2.13)	4	15	12	12	16	13
Критерий α_3 (3.2.2.16)	8	6	21	7	19	14
Критерий Шпигельхальтера (3.2.2.11)	19	13	11	11	8	15
Критерий Саркади (3.2.2.12)	5	18	15	14	13	16
Критерий Смирнова–Крамера–фон Мизеса (3.1.2.2)	17	11	20	17	6	17
Критерий Локка–Спурье (3.2.2.7)	13	4	19	21	17	18
Критерий Оя (3.2.2.8)	20	17	14	13	14	19
Критерий Хегази–Грина (3.2.2.3)	6	19	16	19	21	20
Критерий Муроты–Такеучи (3.2.2.17)	15	21	17	20	20	21

3.3. Критерии проверки экспоненциальности распределения

Экспоненциальный закон распределения вероятностей является базовым законом, используемым в теории надежности. Его аналитическая простота делает его привлекательным для инженеров и исследователей. Однако всегда следует предварительно убедиться в том, что вероятностное поведение случайной величины (например, моментов отказов изделий) подчиняется „желательному" экспоненциальному закону. В ином случае выигрыш от простоты расчетов будет многократно „скомпенсирован" потерями от ошибочных выводов и заключений, вызванных отклонением реального распределения вероятностей случайной величины от экспоненциального закона.

По-видимому, желанием предостеречь пользователя от таких ошибок объясняется многообразие разработанных на сегодняшний день математиками-статистиками критериев экспоненциальности, основные из которых представлены в настоящем разделе.

3.3.1. Критерий Шапиро–Уилка

Критерий предложен Шапиро и Уилком [298]. Предположим, имеется выборка $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$, причем начальная точка распределения неизвестна, т. е. рассматривается плотность вероятностей

$$f(x) = \frac{1}{\nu} \exp\left(\frac{x-\mu}{\nu}\right)$$

с неизвестным параметром μ.

Тогда статистика критерия имеет вид

$$W_E = \frac{n\,(\bar{x} - x_1)^2}{(n-1)\sum(x_i - \bar{x})^2},$$

или при $n \to \infty \quad (n > 50)$

$$W_E = \frac{(\bar{x} - x_1)^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}.$$

Гипотеза экспоненциальности наблюдаемого распределения не отклоняется с достоверностью α, если $W_1(\alpha) \leqslant W_E \leqslant W_2(\alpha)$, где $W_1(\alpha)$ и $W_2(\alpha)$ — критические значения, приведенные в табл. 81.

Для случая цензурированной выборки, когда отсутствуют r_1 наименьших и r_2 наибольших членов выборки, модификация критерия Шапиро–Уилка рассмотрена в [299]. Его статистика в этом случае имеет вид

$$W_1 = \frac{\left(\sum\limits_{i=2}^{n-r_1-r_2} T_{r_1+i}\right)^2}{(n - r_1 - r_2)\sum\limits_{i=2}^{n-r_1-r_2} a_{ij}^{(n-r_1-r_2)} T_{r_1+i} T_{r_1+j}},$$

где

$$T_i = (n-i+1)(x_i - x_{i-1}), \quad i = 2, \ldots, n;$$

$$a_{ij}^{(n)} = \frac{i-1}{n-j-1}, \quad \left(n \geqslant i \geqslant j \geqslant 2;\ a_{ij}^{(n)} = a_{ji}^{(n)};\ i,j = 2,3,\ldots,n\right).$$

Таблица 81

Критические значения
критерия экспоненциальности W_E Шапиро–Уилка [13, 298]

n	$\alpha = 0{,}90$		$\alpha = 0{,}95$		n	$\alpha = 0{,}90$		$\alpha = 0{,}95$	
	$W_1(\alpha)$	$W_2(\alpha)$	$W_1(\alpha)$	$W_2(\alpha)$		$W_1(\alpha)$	$W_2(\alpha)$	$W_1(\alpha)$	$W_2(\alpha)$
7	0,071	0,358	0,062	0,404	22	0,026	0,084	0,023	0,094
8	0,062	0,301	0,054	0,342	23	0,025	0,078	0,022	0,087
9	0,058	0,261	0,050	0,301	24	0,024	0,074	0,021	0,082
10	0,056	0,231	0,049	0,261	25	0,023	0,070	0,021	0,078
11	0,052	0,208	0,046	0,234	26	0,022	0,066	0,020	0,073
12	0,050	0,191	0,044	0,215	27	0,022	0,063	0,020	0,070
13	0,046	0,173	0,040	0,195	28	0,021	0,061	0,019	0,067
14	0,043	0,159	0,038	0,178	29	0,021	0,058	0,019	0,064
15	0,040	0,145	0,036	0,163	30	0,020	0,054	0,018	0,060
16	0,038	0,134	0,034	0,150	31	0,019	0,052	0,017	0,057
17	0,034	0,120	0,030	0,135	32	0,019	0,050	0,017	0,055
18	0,031	0,109	0,028	0,123	33	0,018	0,048	0,017	0,053
19	0,029	0,102	0,026	0,114	34	0,018	0,047	0,017	0,051
20	0,028	0,095	0,025	0,106	35	0,018	0,045	0,016	0,049
21	0,027	0,091	0,024	0,101					

Критические значения статистики W_1 находятся из табл. 81 с заменой n на $n - r_1 - r_2$, т. е. гипотеза экспоненциальности не отклоняется, если

$$W_{1(n-r_1-r_2)}(\alpha) \leqslant W_1 \leqslant W_{2(n-r_1-r_2)}(\alpha).$$

Если начальная точка известна (предположим, $\mu = 0$, что всегда можно сделать заменой x_i на $x_i - \mu$), то статистика Шапиро–Уилка имеет вид

$$W_{E_0} = \frac{\sum (x_i - \bar{x})^2}{\left(\sum x_i\right)^2}.$$

Гипотеза экспоненциальности не отклоняется, если

$$\tilde{W}_1(\alpha) \leqslant W_{E_0} \leqslant \tilde{W}_2(\alpha)\,,$$

где $\tilde{W}_1(\alpha)$, $\tilde{W}_2(\alpha)$ — критические значения, приведенные в табл. 82.

Вместо статистики W_{E_0} можно воспользоваться статистикой [300]

$$\tilde{W}_{E_0} = \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n\left[(n+1)\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2\right]},$$

критические значения которой совпадают с $W_1(\alpha)$ и $W_2(\alpha)$ (см. табл. 81, в которой n следует заменить на $(n+1)$).

Для случая цезурирования справа статистика заменяется на

$$W_2 = \frac{\left(\sum\limits_{i=2}^{n-r_2+1} T_{i-1}\right)^2}{(n-r_2)\sum\limits_{i=2}^{n-r_2+1}\sum\limits_{j=2}^{n-r_2+1} a_{ij}^{(n-r_2+1)} T_{i-1} T_{j-1}}.$$

Таблица 82

Критические значения критерия экспоненциальности W_{E_0} Шапиро–Уилка [13, 298]

n	$\alpha = 0{,}90$		$\alpha = 0{,}95$		n	$\alpha = 0{,}90$		$\alpha = 0{,}95$	
	$\tilde W_1(\alpha)$	$\tilde W_2(\alpha)$	$\tilde W_1(\alpha)$	$\tilde W_2(\alpha)$		$\tilde W_1(\alpha)$	$\tilde W_2(\alpha)$	$\tilde W_1(\alpha)$	$\tilde W_2(\alpha)$
7	0,033	0,225	0,025	0,260	22	0,022	0,069	0,020	0,080
8	0,032	0,200	0,025	0,230	23	0,021	0,065	0,019	0,075
9	0,031	0,177	0,025	0,205	24	0,021	0,062	0,019	0,069
10	0,030	0,159	0,025	0,184	25	0,020	0,058	0,018	0,065
11	0,030	0,145	0,025	0,166	26	0,020	0,056	0,018	0,062
12	0,029	0,134	0,025	0,153	27	0,020	0,054	0,017	0,058
13	0,028	0,124	0,025	0,140	28	0,019	0,052	0,017	0,056
14	0,027	0,115	0,024	0,128	29	0,019	0,050	0,016	0,054
15	0,026	0,106	0,024	0,119	30	0,019	0,048	0,016	0,053
16	0,025	0,098	0,023	0,113	31	0,018	0,047	0,016	0,051
17	0,024	0,093	0,023	0,107	32	0,018	0,045	0,015	0,050
18	0,024	0,087	0,022	0,101	33	0,018	0,044	0,015	0,048
19	0,023	0,083	0,022	0,096	34	0,017	0,043	0,014	0,046
20	0,023	0,077	0,021	0,090	35	0,017	0,041	0,014	0,045
21	0,022	0,074	0,020	0,085					

Ее критические значения совпадают с критическими значениями статистики W_{E_0} (см. табл. 82 с заменой n на $(n - r_2 + 1)$).

Задача 125. *Имеется ряд наблюдений*

$$x_i\colon 1,\ 2,\ 4,\ 5,\ 9,\ 11,\ 18,\ 21,\ 29,\ 35.$$

Проверить соответствие распределения вероятностей случайной величины, представленной этой выборкой, экспоненциальному распределению при достоверности $\alpha = 0{,}95$. Рассмотреть случаи неизвестного начала распределения (μ неизвестно) и известного значения $\mu = 1$.

Случай 1 (μ неизвестно)

Вычисляем $\bar x = 13{,}5$; $(\bar x - x_1)^2 = (3{,}5 - 1)^2 = 156{,}25$; $\sum(x_i - \bar x) = 1256{,}5$.

Тогда $W_E = \dfrac{10 \cdot 156{,}25}{9 \cdot 1256{,}5} = 0{,}138$. Из табл. 81 для $\alpha = 0{,}95$ и $n = 10$ находим $W_1(0{,}95) = 0{,}049$ и $W_2(0{,}95) = 0{,}261$.

Так как $W_1(0{,}95) < W_E < W_2(0{,}95)$, гипотеза экспоненциальности распределения не отклоняется.

Случай 2 ($\mu = 1$)

Находим эквивалентный ряд $x_i - \mu = x_i - 1$: 0, 1, 3, 4, 8, 10, 17, 20, 28, 34, для которого имеем $\bar x = 12{,}5$; $\sum(x_i - \bar x)^2 = 1256{,}5$ и $(\sum x_i)^2 = 156{,}25$.

Тогда $W_{E_0} = \dfrac{1256{,}5}{15625} = 0{,}08$. Из табл. 82 для $\alpha = 0{,}95$ и $n = 10$ находим $\tilde W_1(\alpha) = 0{,}025$ и $\tilde W_2(\alpha) = 0{,}184$.

Так как $\tilde W_1(0{,}95) = 0{,}025 < W_{E_0} = 0{,}08 < \tilde W_2(0{,}95) = 0{,}184$, гипотеза экспоненциальности не отклоняется.

Вычисляем статистику $\tilde W_{E_0} = \dfrac{15625}{10 \cdot (11 \cdot 2819 - 15625)} = 0{,}101$.

Из табл. 81 для $n + 1 = 11$ и $\alpha = 0{,}95$ имеем $W_1(0{,}95) = 0{,}046$ и $W_2(0{,}95) = 0{,}234$, что также не позволяет отклонить гипотезу экспоненциальности, так как $\tilde W_{E_0}$ не выходит за пределы критического диапазона.

Задача 126. *Проверить гипотезу экспоненциальности для ряда наблюдений, представленного в задаче* 125, *при условии, что первое и два последних наблюдения цензурированы* (*для случая, когда* μ *неизвестно и* $\alpha = 0{,}95$).

Имеем $r_1 = 1$ и $r_2 = 2$ $(n - r_1 - r_2 = 7)$. Находим

$$T_3 = 8 \cdot (x_3 - x_2) = 8 \cdot 2 = 16; \quad T_4 = 7 \cdot (x_4 - x_3) = 7; \quad T_5 = 6 \cdot (x_5 - x_4) = 24;$$
$$T_6 = 5 \cdot (x_6 - x_5) = 10; \quad T_7 = 4 \cdot (x_7 - x_6) = 28; \quad T_8 = 3 \cdot (x_8 - x_7) = 9.$$

Вычисляем

$$a_{ij}^{(n-r_1-r_2)} = a_{ij}^{(7)} = \frac{i-1}{7-j+1} = \frac{i-1}{8-j};$$

$$a_{22}^{(7)} = \frac{1}{6}; \quad a_{32}^{(7)} = \frac{2}{6}; \quad a_{42}^{(7)} = \frac{3}{6}; \quad a_{52}^{(7)} = \frac{4}{6}; \quad a_{62}^{(7)} = \frac{5}{6}; \quad a_{72}^{(7)} = \frac{6}{6}; \quad a_{33}^{(7)} = \frac{2}{5};$$
$$a_{43}^{(7)} = \frac{3}{5}; \quad a_{53}^{(7)} = \frac{4}{5}; \quad a_{63}^{(7)} = \frac{5}{5}; \quad a_{73}^{(7)} = \frac{6}{5}; \quad a_{44}^{(7)} = \frac{3}{4}; \quad a_{54}^{(7)} = \frac{4}{4}; \quad a_{64}^{(7)} = \frac{5}{4};$$
$$a_{74}^{(7)} = \frac{6}{4}; \quad a_{55}^{(7)} = \frac{4}{3}; \quad a_{65}^{(7)} = \frac{5}{3}; \quad a_{75}^{(7)} = \frac{6}{3}; \quad a_{66}^{(7)} = \frac{5}{2}; \quad a_{76}^{(7)} = \frac{6}{2}; \quad a_{77}^{(7)} = \frac{6}{1}.$$

Далее получаем

$$6 \cdot \left(\frac{1}{6} \cdot T_3 \cdot T_3 + \frac{2}{6} \cdot T_3 \cdot T_4 + \frac{3}{3} \cdot T_3 \cdot T_5 + \ldots + \frac{5}{3} \cdot T_6 \cdot T_7 + \frac{5}{2} \cdot T_7 \cdot T_7 + \frac{6}{2} \cdot T_7 \cdot T_8\right) =$$
$$= 6 \cdot \left(\frac{16 \cdot 16}{2} + \frac{2 \cdot 16 \cdot 17}{6} + \ldots + \frac{5 \cdot 28 \cdot 28}{2} + \frac{6 \cdot 29 \cdot 9}{2}\right) = 40056.$$

Находим $(T_3 + T_4 + \ldots + T_7 + T_8)_2 = (16 + 7 + \ldots + 28 + 9)^2 = 8836$.

Окончательно получаем $W_1 = \dfrac{8836}{40056} = 0{,}220$. Из табл. 81 для $n - r_1 - r_2 = 7$ и $\alpha = 0{,}95$ находим $W_1(0{,}95) = 0{,}062$ и $W_2(0{,}95) = 0{,}404$.

Так как $W_1(0{,}95) = 0{,}062 < W_E = 0{,}220 < W_2(0{,}95) = 0{,}404$, гипотеза экспоненциальности распределения вероятностей случайных величин не отклоняется.

3.3.2. Критерии типа Колмогорова–Смирнова

Модификации известных критериев типа Колмогорова–Смирнова для применения их при проверке экспоненциальности закона распределения вероятностей (с неизвестными параметрами) наиболее полно представлены в [301, 302].

Предположим, имеет место гипотетический закон распределения вероятностей $F(x) = 1 - \exp\left(-\dfrac{x-\mu}{\nu}\right)$, где μ и ν — неизвестные параметры, оценки которых по выборке могут быть найдены из формул (отметим, что выборка упорядочена, т. е. $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$)

$$\hat{\nu} = \frac{n(\bar{x} - x_1)}{n-1}; \quad \hat{\mu} = x_1 - \frac{\hat{\nu}}{n}.$$

Обозначив $w_i = \dfrac{x_i - \hat{\mu}}{\hat{\nu}}$, переходим к нормированному экспоненциальному распределению $z_i = 1 - \exp(-w_i)$, для которого имеем аналоги известных статистик критериев согласия

— критерий Колмогорова–Смирнова (см. раздел 3.1.2.1):

$$D_n^+ = \max_i \left[\frac{i}{n} - z_i\right]; \quad D_n^- = \max_i \left[z_i - \frac{i-1}{n}\right]; \quad D_n = \max\left(D_n^+, D_n^-\right);$$

— критерий Андерсона–Дарлинга (см. раздел 3.1.2.4):

$$A^2 = -\frac{1}{n} \sum_i (2i-1)[\ln z_i + \ln(1 - z_{n+1-i})] - n;$$

— критерий Смирнова–Крамера–фон Мизеса (см. раздел 3.1.2.2):

$$W^2 = \sum_i \left(z_i - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n};$$

— критерий Ватсона (см. раздел 3.1.2.5):

$$U^2 = W^2 - n\left(\bar{z} - \frac{1}{2}\right)^2, \quad \text{где} \quad \bar{z} = \frac{1}{n}\sum_i z_i;$$

— критерий Купера (см. раздел 3.1.2.6):

$$V = D_n^+ + D_n^-.$$

Для случая проверки экспоненциальности распределения с неизвестными параметрами критические значения перечисленных статистик для различных уровней значимости приведены в табл. 83.

Таблица 83

Критические значения статистик критериев согласия типа Колмогорова–Смирнова для проверки экспоненциальности распределения с неизвестными параметрами [301]

n	Уровень значимости α (верхние процентные точки)					
	0,25	0,15	0,10	0,05	0,025	0,01
	Статистика $\sqrt{n}D_n$					
5	0,683	0,749	0,793	0,865	0,921	0,992
10	0,753	0,833	0,889	0,977	1,048	1,119
15	0,771	0,865	0,912	1,002	1,079	1,163
20	0,786	0,872	0,927	1,021	1,099	1,198
25	0,792	0,878	0,936	1,033	1,115	1,215
50	0,813	0,879	0,960	1,061	1,149	1,257
100	0,824	0,911	0,972	1,072	1,171	1,278
∞	0,840	0,927	0,995	1,094	1,184	1,298
	Статистика V					
5	1,098	1,186	1,234	1,314	1,400	1,494
10	1,194	1,294	1,363	1,461	1,556	1,662
15	1,225	1,325	1,392	1,504	1,596	1,701
20	1,245	1,346	1,419	1,536	1,635	1,769
25	1,260	1,366	1,438	1,559	1,658	1,796
50	1,292	1,400	1,481	1,600	1,701	1,847
100	1,310	1,419	1,502	1,647	1,740	1,897
∞	1,334	1,444	1,532	1,656	1,770	1,910
	Статистика W^2					
5	0,083	0,102	0,117	0,141	0,166	0,197
10	0,097	0,122	0,142	0,176	0,211	0,259
15	0,103	0,130	0,151	0,188	0,229	0,281
20	0,106	0,133	0,157	0,195	0,237	0,293
25	0,107	0,135	0,160	0,199	0,247	0,301
50	0,111	0,141	0,166	0,209	0,256	0,319
100	0,113	0,144	0,170	0,215	0,263	0,328
∞	0,116	0,148	0,175	0,222	0,271	0,338

Окончание таблицы 83

n	Уровень значимости α (верхние процентные точки)					
	0,25	0,15	0,10	0,05	0,025	0,01
	Статистика U^2					
5	0,068	0,083	0,093	0,113	0,131	0,153
10	0,075	0,094	0,108	0,131	0,155	0,187
15	0,080	0,099	0,114	0,139	0,165	0,200
20	0,082	0,102	0,117	0,143	0,170	0,207
25	0,083	0,104	0,119	0,146	0,173	0,212
50	0,087	0,108	0,124	0,152	0,180	0,223
100	0,089	0,110	0,126	0,155	0,184	0,229
∞	0,090	0,112	0,129	0,159	0,189	0,236
	Статистика A^2					
5	0,460	0,555	0,621	0,725	0,848	0,989
10	0,545	0,660	0,747	0,920	1,068	1,352
15	0,575	0,720	0,816	1,009	1,198	1,495
20	0,608	0,757	0,861	1,062	1,267	1,580
25	0,625	0,784	0,890	1,097	1,317	1,635
50	0,680	0,838	0,965	1,197	1,440	1,775
100	0,710	0,875	1,008	1,250	1,510	1,865
∞	0,736	0,910	1,062	1,321	1,591	1,959

Для $n \geqslant 5$ верхние процентные точки распределения статистик вполне удовлетворительно совпадают с предельными распределениями при использовании модификаций, приведенных в табл. 84.

Таблица 84

Процентные точки модифицированных критериев типа Колмогорова–Смирнова для проверки экспоненциальности распределения

Статистика	Модификация	Уровень значимости α					
		0,25	0,15	0,10	0,05	0,025	0,01
W^2	$W^2\left(1+\dfrac{2{,}8}{n}-\dfrac{3}{n^2}\right)$	0,116	0,148	0,175	0,222	0,271	0,338
U^2	$U^2\left(1+\dfrac{2{,}3}{n}-\dfrac{3}{n^2}\right)$	0,090	0,112	0,129	0,159	0,189	0,230
A^2	$A^2\left(1+\dfrac{5{,}4}{n}-\dfrac{11}{n^2}\right)$	0,736	0,916	1,062	1,321	1,591	1,959

Задача 127. *Проверить гипотезу экспоненциальности распределения вероятностей случайных величин в условиях задачи* 125 *на уровне значимости* $\alpha = 0{,}05$ *критериями типа Колмогорова–Смирнова.*

Имеем $\bar{x} = 13{,}5$; $\hat{\nu} = \dfrac{10\cdot(13{,}5-1)}{9} = 13{,}889$; $\hat{\mu} = 1 - \dfrac{13{,}889}{10} = -0{,}3889$.

Далее вычисляем

$$W_1 = \frac{x_1-\hat{\mu}}{\hat{\nu}} = \frac{1+0{,}3889}{13{,}889} = 0{,}1; \quad z_i = 1 - e^{-0{,}1} = 0{,}0952;$$

$W_2 = 0{,}172$; $z_2 = 0{,}1580$; $W_3 = 0{,}136$; $z_3 = 0{,}2709$; $W_4 = 0{,}388$; $z_4 = 0{,}3216$;
$W_5 = 0{,}676$; $z_5 = 0{,}4913$; $W_6 = 0{,}820$; $z_6 = 0{,}5596$; $W_7 = 1{,}324$; $z_7 = 0{,}7339$;
$W_8 = 1{,}540$; $z_8 = 0{,}7856$; $W_9 = 2{,}116$; $z_9 = 0{,}8795$; $W_{10} = 2{,}548$; $z_{10} = 0{,}9218$.

Критерий Колмогорова–Смирнова

Результаты расчетов представлены в таблице:

i	$\frac{i}{n}$	z_i	$\frac{i}{n} - z_i$	$\frac{i-1}{n}$	$z_i - \frac{i-1}{n}$
1	0,1	0,0952	0,0048	0,0	0,0952
2	0,2	0,1580	0,0420	0,1	0,0580
3	0,3	0,2709	0,0291	0,2	0,0709
4	0,4	0,3216	0,0784	0,3	0,0216
5	0,5	0,4913	0,0087	0,4	0,0913
6	0,6	0,5596	0,0404	0,5	0,0956
7	0,7	0,7339	−0,0339	0,6	0,1339
8	0,8	0,7856	0,0144	0,7	0,0856
9	0,9	0,8795	0,0205	0,8	0,0795
10	1,0	0,9218	0,0782	0,9	0,0218

Из таблицы видим, что

$$D_n^+ \max\left[\frac{i}{n} - z_i\right] = 0{,}0784; \quad D_n^- = \max\left[z_i - \frac{i-1}{n}\right] = 0{,}1339;$$

$$D_n = \max\left(D_n^+, D_n^-\right) = 0{,}1339.$$

Значение статистики критерия равно $\sqrt{n} \cdot D_n = \sqrt{10} \cdot 0{,}1339 = 0{,}423$.

Из табл. 83 для $\alpha = 0{,}05$ и $n = 10$ находим, что критическое значение статистики равно 0,97. Так как $\sqrt{n} \cdot D_n = 0{,}423 < 0{,}97$, гипотеза экспоненциальности распределения вероятностей случайных величин не отклоняется.

Критерий Купера

Статистика критерия равна

$$\sqrt{n} \cdot V_n = \sqrt{n} \cdot (D_n^+ + D_n^-) = \sqrt{10} \cdot (0{,}0784 + 0{,}1339) = 0{,}671.$$

Из табл. 83 для $\alpha = 0{,}05$ и $n = 10$ находим критическое значение, равное 1,461.

Так как $\sqrt{n} \cdot V_n = 0{,}671 < 1{,}461$, гипотеза экспоненциальности не отклоняется.

Критерий Андерсона–Дарлинга

Вычисляем

$$A^2 = -\frac{1}{10} \cdot \Big\{1 \cdot [\ln 0{,}0952 + \ln(1 - 0{,}9218)] + 3 \cdot [\ln 0{,}158 + \ln(1 - 0{,}8795)] + \\ + 5 \cdot [\ln 0{,}2709 + \ln(1 - 0{,}8756)] + 7 \cdot [\ln 0{,}3216 + \ln(1 - 0{,}7339)] + \\ + 9 \cdot [\ln 0{,}4913 + \ln(1 - 0{,}3216)] + 11 \cdot [\ln 0{,}5596 + \ln(1 - 0{,}4913)] + \\ + 13 \cdot [\ln 0{,}7339 + \ln(1 - 0{,}3216)] + 15 \cdot [\ln 0{,}7856 + \ln(1 - 0{,}2709)] + \\ + 17 \cdot [\ln 0{,}8795 + \ln(1 - 0{,}158)] + 19 \cdot [\ln 0{,}9218 + \ln(1 - 0{,}0952)]\Big\} - 10 = 0{,}17988.$$

Из табл. 83 для $n = 10$ и $\alpha = 0{,}05$ находим критическое значение $A^2(0{,}05) = 0{,}920$ или для модифицированного критерия

$$\left(A^2\right)' = A^2 \cdot \left(1 + \frac{5{,}4}{n} - \frac{11}{n^2}\right) = 0{,}17988 \cdot \left(1 + \frac{5{,}4}{10} - \frac{11}{100}\right) = 0{,}257,$$

критическое значение которого из табл. 84 равно 1,321. Далее имеем

$$A^2 = 0{,}17988 < A^2(0{,}05) = 0{,}920; \quad \left(A^2\right)' = 0{,}257 < \left(A^2(0{,}05)\right)' = 1{,}321,$$

и гипотеза экспоненциальности не отклоняется.

Критерий Смирнова–Крамера–фон Мизеса

Вычисляем статистики критерия

$$W^2 = \left(0{,}0952 - \frac{1}{20}\right)^2 + \left(0{,}158 - \frac{3}{20}\right)^2 + \left(0{,}2709 - \frac{5}{20}\right)^2 + \left(0{,}3216 - \frac{7}{20}\right)^2 + \\ + \left(0{,}4913 - \frac{9}{20}\right)^2 + \left(0{,}5596 - \frac{11}{20}\right)^2 + \left(0{,}7339 - \frac{13}{20}\right)^2 + \left(0{,}7856 - \frac{15}{20}\right) + \\ + \left(0{,}8795 - \frac{17}{20}\right)^2 + \left(0{,}9218 - \frac{19}{20}\right)^2 + \frac{1}{120} = 0{,}0234$$

и $\left(W^2\right)' = W^2 \cdot \left(1 + \frac{2{,}8}{n} - \frac{3}{n^2}\right) = 0{,}0234 \cdot (1 + 0{,}28 - 0{,}03) = 0{,}0295.$

Из табл. 84 имеем критическое значение статистики

$$(W^2(\alpha))' = 0{,}222 \text{ (для } n = 10 \text{ и } \alpha = 0{,}05).$$

Так как $(W^2)' = 0{,}0292 < (W^2(0{,}05))' = 0{,}222$, гипотеза экспоненциальности не отклоняется.

Критерий Ватсона

$$z = \frac{1}{n} \cdot \sum z_i = 0{,}52174;$$

$$U^2 = W^2 - n \cdot \left(\bar{z} - \frac{1}{2}\right)^2 = 0{,}0234 - 10 \cdot (0{,}52174 - 0{,}5)^2 = 0{,}01877;$$

$$\left(U^2\right)' = U^2 \cdot \left(1 + \frac{2{,}3}{n} - \frac{3}{n^2}\right) = 0{,}01877 \cdot (1 + 0m23 - 0{,}03) = 0{,}0255.$$

Из табл. 84 для $n = 10$ и $\alpha = 0{,}05$ имеем критическое значение $(U^2(\alpha))' = 0{,}159$. Так как $(U^2)' = 0{,}0225 < (U^2(\alpha))' = 0{,}159$, гипотеза экспоненциальности распределения не отклоняется.

3.3.3. Критерии типа Смирнова–Крамера–фон Мизеса для цензурированных данных

Рассматривается выборка, в которой все наблюдения, бóльшие, чем некоторая величина x_p, цензурированы, т. е. в распоряжении наблюдателя есть только r наблюдений из выборки объема n, удовлетворяющих условию $0 \leqslant x_1 \leqslant x_2 \leqslant \ldots \leqslant x_p$.

Необходимо проверить гипотезу о том, что ряд наблюдений извлечен из выборки с функцией распределения вероятностей

$$F(x) = 1 - \exp\left(-\frac{x}{\nu}\right).$$

Оценкой ν по выборке является

$$\hat{\nu} = \frac{1}{r}\left[\sum_{i=1}^{n} x_i + (n - r)\, x_p\right].$$

Оценкой степени цезурирования p является величина

$$\hat{p} = 1 - \exp\left(-\frac{x_p}{\hat{\nu}}\right).$$

Применение критерия Смирнова–Крамера–фон Мизеса (см. раздел 3.1.2.2) для цензурированных выборок рассмотрено в [304]. Полагая, что $z_i = 1 - \exp\left(-\frac{x_i}{\nu_i}\right)$,

запишем статистику критерия в виде

$$W_p^2 = \sum_{i=1}^{r}\left(z_i - \frac{2i-1}{2n}\right)^2 + \frac{1}{12r}.$$

Критические значения статистики W_p^2 для выборок со степенью цензурирования p для различных уровней значимости приведены в табл. 85.

Таблица 85

Критические значения статистики W_p^2
(α — уровень значимости) [304]

α	Степень цензурирования p					
	0,5	0,6	0,7	0,8	0,9	1,0
0,15	0,0531	0,0720	0,0921	0,1126	0,1321	0,1480
0,10	0,0635	0,0857	0,1093	0,1333	0,1561	0,1745
0,05	0,0821	0,1103	0,1401	0,1702	0,1986	0,2216
0,025	0,1015	0,1359	0,1721	0,2087	0,2433	0,2706
0,01	0,1279	0,1710	0,2160	0,2613	0,3033	0,3376

Рекомендуется использовать модифицированную форму критерия

$$\left(W_p^2\right)' = W_p^2\left(1 + \frac{2{,}8}{r} - \frac{3}{r^2}\right).$$

Естественно, при $p = 1$ (отсутствие цензурирования) критические значения статистики совпадают с приведенными в табл. 84.

Задача 128. *Проверить гипотезу экспоненциальности распределения случайных величин в условиях задачи* 125 *на уровне значимости* $\alpha = 0{,}05$ *критериями Смирнова–Крамера–фон Мизеса и Колмогорова–Смирнова при условии, что все наблюдения, бóльшие* $x_p = 20$, *цензурированы.*

Имеем $r = 7$ и $x_p = 20$. Вычисляем

$$\hat{\nu} = \frac{1}{7}\cdot\left[\sum_{i=1}^{7} x_i + (10-7)\cdot 20\right] = 15{,}71428; \quad \hat{p} = 1 - \exp\left(-\frac{20}{15{,}71428}\right) = 0{,}72;$$

$$z_1 = 1 - \exp\left(-\frac{1}{15{,}71428}\right) = 0{,}06165; \quad z_2 = 0{,}1195; \quad z_3 = 0{,}22473;$$

$$z_4 = 0{,}2725; \quad z_5 = 0{,}4360; \quad z_6 = 0{,}5034; \quad z_7 = 0{,}6819.$$

Далее

$$W_p^2 = \left(0{,}06165 - \frac{1}{20}\right)^2 + \left(0{,}1195 - \frac{3}{20}\right)^2 + \left(0{,}22473 - \frac{5}{20}\right)^2 + \left(0{,}2775 - \frac{7}{20}\right)^2 +$$
$$+ \left(0{,}4360 - \frac{9}{20}\right)^2 + \left(0{,}5034 - \frac{11}{20}\right) + \left(0{,}6819 - \frac{13}{20}\right)^2 + \frac{1}{12\cdot 7} = 0{,}023$$

и

$$\left(W_p^2\right)' = 0{,}023\cdot\left(1 + \frac{2{,}8}{7} - \frac{3}{49}\right) = 0{,}0308.$$

Из табл. 85 для $\alpha = 0{,}05$ и $p = 0{,}72$ экстраполяцией получаем $(W_p^2(0{,}05))' = 0{,}15$. Так как $(W_p^2)' = 0{,}0308 < (W_p^2(\alpha))' = 0{,}15$, гипотеза экспоненциальности не отклоняется.

3.3.4. Критерий Фроцини

В [239] рассмотрен критерий экспоненциальности, основанный на статистике

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| 1 - \exp\left(-\frac{x_i}{\bar{x}} \right) - \frac{i - 0{,}5}{n} \right|,$$

критические значения которой $B_n(\alpha)$ приведены в табл. 86.

Таблица 86

Критические значения $B_n(\alpha)$ критерия экспоненциальности Фроцини (α — доверительная вероятность)

n	α			n	α		
	0,90	0,95	0,99		0,90	0,95	0,99
5	0,3261	0,3687	0,4499	14	0,3373	0,3821	0,4656
6	0,3241	0,3666	0,4495	15	0,3364	0,3837	0,4747
7	0,3292	0,3742	0,4584	16	0,3345	0,3777	0,4693
8	0,3289	0,3740	0,4609	17	0,3387	0,3806	0,4716
9	0,3365	0,3800	0,4660	18	0,3360	0,3814	0,4730
10	0,3377	0,3820	0,4753	19	0,3370	0,3844	0,4796
11	0,3334	0,3790	0,4710	20	0,3351	0,3795	0,4738
12	0,3318	0,3784	0,4641	∞	0,3380	0,3840	0,4760
13	0,3313	0,3768	0,4631				

Мощность критерия не уступает всем известным до $n \leqslant 20$ и превосходит их при $n > 50$.

Задача 129. *В условиях задачи* 125 *проверить гипотезу экспоненциальности распределения вероятностей случайных величин критерием Фроцини при доверительной вероятности* $\alpha = 0{,}95$.

Находим при $\bar{x} = 13{,}5$

$$B = \frac{1}{\sqrt{10}} \cdot \left(\left| 1 - \exp\left(-\frac{1}{13{,}5} \right) - \frac{0{,}5}{10} \right| + \left| 1 - \exp\left(-\frac{2}{13{,}5} \right) - \frac{1{,}5}{10} \right| + \ldots \right.$$
$$\left. \ldots + \left| 1 - \exp\left(-\frac{9{,}5}{13{,}5} \right) - \frac{9{,}5}{10} \right| \right) = 0{,}0974.$$

Из табл. 86 для $n = 10$ находим $B_n(0{,}95) = 0{,}38$.

Так как $B_n = 0{,}097 < B_n(0{,}95) = 0{,}38$, гипотеза экспоненциальности распределения не отклоняется.

3.3.5. Корреляционный критерий экспоненциальности

Рассмотрим функцию распределения

$$F(x) = 1 - \exp\left(-\frac{x - \mu}{\nu} \right),$$

параметры которой оцениваются по формулам

$$\hat{\nu} = \frac{m(\bar{x} - x_1)}{n - 1} \quad \text{и} \quad \hat{\mu} = x_1 - \frac{\hat{\nu}}{n}, \quad \left(\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \right).$$

В [301] рассмотрен критерий экспоненциальности, аналогичный корреляционному критерию Филлибена проверки нормальности распределения (см. раздел 3.2.2.4).

Статистика критерия основана на коэффициенте корреляции r между нормированной переменной $z_i = \dfrac{x_i - \hat{\mu}}{\hat{\nu}}$ и математическим ожиданием i-й порядковой статистики из экспоненциального распределения, представленного выборкой объема n [303]:

$$m_i = \sum_{j=1}^{i} \frac{1}{n-j+1}, \qquad r(z,m) = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(m_i - \bar{m})}{\left\{\sum_{i=1}^{n}(z_i - \bar{z})^2 \sum_{i=1}^{n}(m_i - \bar{m})^2\right\}^{\frac{1}{2}}}.$$

Если вместо точного значения m_i используется (при $n > 20$) его аппроксимация $\tilde{m}_i = -\ln\left(1 - \dfrac{i}{n+1}\right)$, то соответствующий коэффициент корреляции будем обозначать $r(z,\tilde{m})$. Статистика критерия используется в форме

$$K(z,m) = n\big[1 - r^2(z,m)\big] \quad \text{или} \quad K(z,\tilde{m}) = n\big[1 - r^2(z,\tilde{m})\big].$$

Критические значения статистик приведены в табл. 87.

Таблица 87

Критические значения $K(z,m,\alpha)$ и $K(z,\tilde{m},\alpha)$ корреляционного критерия экспоненциальности
(α — уровень значимости) [303]

n	$K(z,m,\alpha)$			$K(z,\tilde{m},\alpha)$		
	$\alpha = 0{,}10$	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}10$	$\alpha = 0{,}05$	$\alpha = 0{,}01$
5	1,005	1,305	1,960	0,960	1,175	1,760
10	1,560	1,920	2,720	1,540	1,990	2,850
15	1,935	2,445	3,480	1,980	2,550	3,885
20	2,260	2,820	4,200	2,320	3,040	4,820
25	2,425	3,075	4,725	2,550	3,375	5,575
50	3,350	4,250	7,150	3,650	5,050	8,800
100	4,300	5,700	10,500	5,000	6,900	12,900

Задача 130. *Проверить гипотезу экспоненциальности распределения в условиях задачи 125 корреляционным критерием на уровне значимости* $\alpha = 0{,}10$.

Имеем $\hat{\nu} = \dfrac{10 \cdot (13{,}5 - 1)}{9} = 13{,}889$; $\hat{\mu} = 1 - \dfrac{13{,}889}{10} = -0{,}3889$.

Получаем ряд нормированных переменных

z_i: 0,1; 0,172; 0,316; 0,388; 0,676; 0,820; 1,324; 1,540; 2,116; 2,548.

Вычисляем

$$m_1 = \frac{1}{10}; \quad m_2 = \frac{1}{10} + \frac{1}{9} = 0{,}2111; \quad m_3 = 0{,}33611; \quad m_4 = 0{,}47897; \quad m_5 = 0{,}64563;$$
$$m_6 = 0{,}84563; \quad m_7 = 1{,}09563; \quad m_8 = 1{,}42897; \quad m_9 = 1{,}92897; \quad m_{10} = 2{,}92897.$$

Тогда

$$\sum_{i=1}^{n}(z_i - \bar{z})^2 = 6{,}513696; \quad \sum_{i=1}^{n}(m_i - \bar{m}) = 7{,}07104; \quad \sum_{i=1}^{n}(z_i - \bar{z}) = 6{,}66388;$$

$$r(z,m) = 6{,}66388 \cdot (6{,}513696 \cdot 7{,}07104)^{-\frac{1}{2}} = 0{,}98191;$$

$$K(z,m) = 10 \cdot \left(1 - 0{,}98191^2\right) = 0{,}358.$$

Из табл. 87 для $\alpha = 0{,}10$ и $n = 10$ находим критическое значение $K(z, m, \alpha) = 1{,}56$, и так как полученное значение $K(z, m) = 0{,}358$ меньше критического 1,56, гипотеза экспоненциальности не отклоняется.

Если бы мы вместо m_i использовали аппроксимацию $\tilde{m}_i = -\ln\left(1 - \dfrac{i}{n+1}\right)$, то получили бы

$$\tilde{m}_1 = -\ln\left(1 - \frac{1}{11}\right) = 0{,}0953; \quad \tilde{m}_2 = 0{,}20067;$$

$$\tilde{m}_3 = 0{,}31834; \quad \tilde{m}_4 = 0{,}4520; \quad \tilde{m}_5 = 0{,}6061; \quad \tilde{m}_6 = 0{,}7884;$$

$$\tilde{m}_7 = 1{,}0116; \quad \tilde{m}_8 = 1{,}2993; \quad \tilde{m}_9 = 1{,}7047; \quad \tilde{m}_{10} = 2{,}3979;$$

$$\sum_{i=1}^{n}\left(\tilde{m}_i - \bar{\tilde{m}}\right)^2 = 4{,}836; \quad \bar{\tilde{m}} = 0{,}887428; \quad \sum_{i=1}^{n}(z_i - \bar{z}) \cdot \left(\tilde{m}_i - \bar{\tilde{m}}\right) = 5{,}55767;$$

$$r(z, \tilde{m}) = 5{,}55767 \cdot (6{,}513693 \cdot 4{,}836)^{-\frac{1}{2}} = 0{,}99023; \quad K(z, \tilde{m}) = 10 \cdot \left(1 - 0{,}99023^2\right) = 0{,}194,$$

что меньше, чем критическое значение $K(z,\tilde{m},\alpha) = 1{,}99$ (см. табл. 87).

Следовательно, и в этом случае гипотеза экспоненциальности не отклоняется.

3.3.6. Регрессионный критерий Брейна–Шапиро

Если $x_1, \ldots, x_n$ — порядковые статистики из экспоненциального распределения с плотностью

$$f(x) = \lambda \exp\{-\lambda(x - \mu)\},$$

то случайные величины $y_i = (n - i + 1)(x_i - x_{i-1})$ являются независимыми и идентично распределенными экспоненциальными величинами.

В [305] предложен весьма эффективный критерий экспоненциальности, основанный на наклоне линии регрессии взвешенных спейсингов y_i (т. е. разности смежных членов выборки) на их порядковые номера i. Так как интенсивность отказов для экспоненциального распределения $\lambda = \text{const}$, то проверка экспоненциальности эквивалентна проверке равенства нулю угла наклона регрессии $y_i = f(i)$ (о регрессии можно прочитать в главе 5).

Для общего случая цензурированной выборки, когда отсутствуют s наименьших и r наибольших наблюдений $(m = n - r - s)$, статистика критерия имеет вид

$$z = \frac{\displaystyle\sum_{i=1}^{m-1}\left(i - \frac{m}{2}\right)(y_{s+i+1} - \bar{y})}{\displaystyle\sum_{i=1}^{m-1} y_{s+i+1}\left\{\frac{1}{m(m-1)}\sum_{i=1}^{m-1}\left(i - \frac{m}{2}\right)^2\right\}^{\frac{1}{2}}},$$

где $\bar{y} = \dfrac{1}{m-1}\displaystyle\sum_{i=1}^{m-1} y_{s+i+1}$, или в более компактной форме

$$z = \sqrt{\frac{12}{m-2}}\,\frac{\displaystyle\sum_{i=1}^{m-1}\left(i - \frac{m}{2}\right) y_{s+i+1}}{\displaystyle\sum_{i=1}^{m-1} y_{+i+1}}.$$

Если ввести обозначения

$$t_i = \sum_{j=1}^{i} y_{s+i+1}, \ (i = 1, \ldots, m-1); \quad U_i = \frac{t_i}{t_m - 1}, \ (i = 1, \ldots, m-2),$$

то $z = \sqrt{12(m-2)}\left(\bar{U} - \frac{1}{12}\right)$, где $\bar{U} = \frac{1}{m-2}$.

Если $s = 0$ и μ известно, то $m = n - r + 1$ и $t_i = \sum\limits_{j=1}^{i} y_j$ $(j = 1, \ldots, m-1)$, т. е. добавляется спейсинг $n(x_1 - \mu)$.

Гипотеза экспоненциальности отклоняется на уровне значимости α, если $|z| > u_{\frac{\alpha}{2}}$, где $u_{\frac{\alpha}{2}}$ — $\left(1 - \frac{\alpha}{2}\right)$-квантиль стандартного нормального распределения.

z-критерий используется против альтернативы, утверждающей, что интенсивность отказов (параметр λ) монотонно изменяется в выборке.

Для альтернативы, утверждающей, что интенсивность отказов меняется монотонно, предложена статистика

$$\tilde{z} = \left\{\frac{5}{4(m+1)(m-2)(k-3)}\right\}^{\frac{1}{2}} \frac{12\sum\limits_{i=1}^{m-1} a_i^2 y_{s+i+1} - m(m-2)\sum\limits_{i=1}^{m-1} y_{s+i+1}}{\sum\limits_{i=1}^{m-1} y_{s+i+1}},$$

где $a_i = i - \frac{m}{2}$, $i = 1, \ldots, m-1$, или

$$\tilde{z} = \left\{\frac{5(m-2)}{(m+1)(m-3)}\right\}^{\frac{1}{2}} \left\{m - 3 + 6(m-1)\,\bar{U} - \frac{12}{m-2}\sum_{i=1}^{m-2} iU_i\right\}.$$

Статистика имеет стандартное нормальное распределение. Если априорные сведения о возможном характере альтернативы отсутствуют, рекомендуется использовать комплексную статистику $z^* = z^2 + \tilde{z}^2$, критические значения которой при доверительной вероятности α можно вычислить по формулам

$$z^*_\alpha = \begin{cases} 4{,}605 - \dfrac{2{,}5}{m}, & \text{при } \alpha = 0{,}90; \\ 5{,}991 - \dfrac{1{,}25}{m}, & \text{при } \alpha = 0{,}95; \\ 7{,}378 + \dfrac{3{,}333}{m}, & \text{при } \alpha = 0{,}99. \end{cases}$$

Исследования авторов работы [305] позволяют сделать вывод о высокой мощности этого критерия.

Задача 131. *Для данных задачи* 125 *проверить гипотезу экспоненциальности регрессионным критерием Брейна–Шапиро при доверительной вероятности* $\alpha = 0{,}95$. *Известно, что цензурированию подверглось первое* ($x_1 = 11$; $s = 1$) *и два наибольших* ($x_9 = 29$ *и* $x_{10} = 35$) *значения* ($r = 2$).

Имеем $m = 10 - 1 - 2 = 7$. Находим $y_{s+i+1} = y_{2+i}$ и для $i = 1, \ldots, 6$ получаем:

$$y_3 = (10 - 3 + 1)\cdot(x_3 - x_2) = 8\cdot(4-2) = 16; \quad y_4 = (10 - 4 + 1)\cdot(x_4 - x_3) = 7;$$
$$y_5 = 24; \quad y_6 = 10; \quad y_7 = 28; \quad y_8 = 9;$$

$$t_1 = \sum_{j=1}^{1} y_{2+j} = y_3 = 16; \quad t_2 = \sum_{j=1}^{2} y_{2+j} = y_3 + y_4 = 23; \quad t_3 = \sum_{j=1}^{3} y_{2+j} = 23 + y_5 = 47;$$

$$t_4 = 57; \quad t_5 = 85; \quad t_6 = 94;$$

$$U_1 = \frac{t_1}{t_6} = \frac{16}{94} = 0{,}17023; \quad U_2 = \frac{23}{94} = 0{,}2447; \quad U_3 = 0{,}5; \quad U_4 = 0{,}6064; \quad U_5 = 0{,}9042;$$

$$\bar{U} = 0{,}4851.$$

Далее вычисляем

$$z = \{12 \cdot (7-2)\}^{\frac{1}{2}} \cdot (0{,}4851 - 0{,}5) = -0{,}11541; \quad z^2 = 0{,}01332;$$

$$\sum_{i=1}^{5} i \cdot U_i = 0{,}1702 + 2 \cdot 0{,}2447 + \ldots + 5 \cdot 0{,}9042 = 9{,}1062;$$

$$\tilde{z} = \sqrt{\frac{5 \cdot 5}{8 \cdot 4}} \cdot \left(7 - 3 + 6 \cdot 6 \cdot 0{,}4851 - \frac{12}{5} \cdot \sum_{i=1}^{5} i \cdot U_i\right) =$$

$$= 0{,}88388 \cdot (21{,}4636 - 2{,}4 \cdot 9{,}1062) = -0{,}34584; \quad z^* = 0{,}0133 + 0{,}1196 = 0{,}13229.$$

Так как $z^* = 0{,}13229 < z^*_\alpha = 5{,}991 - \dfrac{1{,}25}{7} = 5{,}812$, гипотеза экспоненциальности не отклоняется.

3.3.7. Критерий Кимбера–Мичела

Кимбер [306] рассмотрел еще один критерий, основанный на линейной зависимости между теоретической $F(x)$ и эмпирической $F_n(x) = \dfrac{i}{n}$ функциями распределения вероятностей случайных величин. Для того, чтобы стабилизировать эту зависимость и ослабить влияние неравных дисперсий $F(x)$ и $F_n(x)$, Мичел [307] предложил стабилизирующие преобразования

$$s_i = \frac{2}{\pi} \arcsin \sqrt{F(x_i)}; \quad r_i = \frac{2}{\pi} \sqrt{\frac{i - 0{,}5}{n}}.$$

При $n \to \infty$ и $\dfrac{i}{n} \to p$ асимптотическая дисперсия $\dfrac{\sqrt{n}}{s_i} \to \dfrac{1}{\pi^2}$, т. е. не зависит от p. Статистика критерия имеет вид

$$D = \max_i |r_i - s_i|,$$

в т. ч. для случая $F(z) = 1 - \exp(-z)$, где $z = \dfrac{x}{\hat{\nu}}$ — стандартизированная случайная экспоненциальная величина $\left(\hat{\nu} = \dfrac{1}{n} \sum x_i\right)$. Критические значения D-статистики приведены в табл. 88.

Таблица 88

Критические значения $D(\alpha)$ статистики Кимбера–Мичела [307]

n	Уровень значимости α			n	Уровень значимости α		
	0,10	0,05	0,01		0,10	0,05	0,01
3	0,240	0,265	0,301	20	0,147	0,163	0,199
4	0,232	0,265	0,310	30	0,126	0,140	0,171
5	0,222	0,251	0,301	40	0,113	0,125	0,152
6	0,216	0,244	0,296	60	0,096	0,106	0,129
8	0,198	0,223	0,272	80	0,085	0,094	0,114
10	0,185	0,208	0,253	100	0,077	0,085	0,103
14	0,166	0,185	0,226				

Мощность этого критерия для многих альтернатив выше, чем мощность критериев Колмогорова–Смирнова (см. раздел 3.3.2) и Шапиро–Уилка (см. раздел 3.3.1). Применение критерия подобного типа для построения критерия согласия с двухпараметрическим распределением Вейбулла рассмотрено в [308].

Задача 132. *Проверить гипотезу экспоненциальности распределения случайных величин для данных задачи* 125 *критерием Кимбера–Мичела на уровне значимости* $\alpha = 0{,}05$.

Находим

$$s_1 = \frac{2}{\pi} \cdot \arcsin\left[1 - \exp\left(-\frac{1}{13{,}5}\right)\right]^{\frac{1}{2}} = 0{,}172198;$$

$$s_2 = 0{,}24202; \quad s_3 = 0{,}33804; \quad s_4 = 0{,}37559; \quad s_5 = 0{,}49146;$$
$$s_7 = 0{,}65676; \quad s_8 = 0{,}69611; \quad s_9 = 0{,}77805; \quad s_{10} = 0{,}823609;$$

$$r_1 = \frac{2}{\pi} \cdot \arcsin(0{,}05)^{\frac{1}{2}} = 0{,}143564; \quad r_2 = 0{,}253183; \quad r_3 = 0{,}3333; \quad r_4 = 0{,}4030;$$
$$r_5 = 0{,}4681; \quad r_6 = 0{,}5319; \quad r_7 = 0{,}5970; \quad r_8 = 0{,}6666; \quad r_9 = 0{,}7468; \quad r_{10} = 0{,}8564.$$

Получаем ряд значений

$$|r_i - s_i|: \ 0{,}0286; \ 0{,}0112; \ 0{,}00471; \ 0{,}0274; \ 0{,}0404; \ 0{,}00464; \ 0{,}05976; \ 0{,}0295; \ 0{,}03125; \ 0{,}03279.$$

Тогда $D = \max\limits_i |r_i - s_i| = 0{,}05976$. Из табл. 88 для $n = 10$ и $\alpha = 0{,}05$ находим $D(\alpha) = 0{,}208$. Так как $D = 0{,}05976 < D(\alpha) = 0{,}208$, гипотеза экспоненциальности не отклоняется.

3.3.8. Критерий Фишера

Если мы имеем ряд $x_1, x_2, \ldots, x_n$, то статистика критерия имеет вид

$$F = \frac{\sum\limits_{i=1}^{n} x_i}{(n-1)\, x_1}.$$

Эта статистика при справедливости нулевой гипотезы (т. е. экспоненциальности распределения) имеет F-распределение Фишера с $f_1 = 2n - 2$ и $f_2 = 2$ степенями свободы. Если

$$\frac{\sum\limits_{i=1}^{n} x_i}{(n-1)\, x_1} > F_\alpha(2n-2, 2),$$

где $F_\alpha(f_1, f_2)$ — $\alpha \cdot 100\%$-е критическое значение F-статистики с f_1 и f_2 степенями свободы, то нулевая гипотеза экспоненциальности отклоняется.

Критические значения F-статистики могут быть определены по таблицам распределения Фишера, например, в [24, 25, 29, 57] или с помощью аппроксимаций, приведенных в разделе 1.1.10.

Задача 133. *Проверить гипотезу экспоненциальности критерием Фишера для данных задачи* 125 *при уровне достоверности* $\alpha = 0{,}95$.

Имеем $n = 10$, $x_1 = 1$ и $\sum x_i = 135$, $F = \frac{135}{9} = 15$. Из таблиц распределения Фишера [25] для $F_{0,95}(18, 2)$ находим $F_{0,95}(18{,}2) = 19{,}3$.

Так как $F = 15 < F_{0,95}(18{,}2) = 19{,}3$, гипотеза экспоненциальности не отклоняется.

3.3.9. Критерий Бартлетта–Морана

Статистика критерия, вычисляемая по ряду значений $x_1, x_2, \ldots, x_n$, имеет вид [309, 310]

$$B = \frac{2n\left[\ln\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + \frac{1}{n}\sum_{i=1}^{n}\ln x_i\right]}{1 + \frac{n+1}{6n}} = \frac{12n^2}{7n+1}\left[\ln\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + \frac{1}{n}\sum_{i=1}^{n}\ln x_i\right].$$

При $n \geqslant 20$ распределение статистики B удовлетворительно аппроксимируется χ^2-распределением с $f = n-1$ степенями свободы [95]. Поэтому нулевая гипотеза экспоненциальности отклоняется, если $B > \chi^2_\alpha(n-1)$ на уровне значимости $(1-\alpha)$.

Задача 134. *В условиях задачи 125 проверить гипотезу экспоненциальности критерием Бартлетта–Морана при $\alpha = 0{,}95$.*

Имеем

$$\bar{x} = 13{,}5; \quad \ln\bar{x} = 2{,}6027; \quad \frac{1}{n}\cdot\sum_{i=1}^{n}\ln x_i = 2{,}1141; \quad B = \frac{2\cdot 100}{70+1}\cdot(2{,}6027 + 2{,}1141) = 3{,}7867.$$

Из табл. 55 для $f = n-1 = 9$ и $\alpha = 0{,}95$ находим $\chi^2_{0,95}(9) = 16{,}919$.

Так как $B = 13{,}2867 < \chi^2_{0,95}(9) = 16{,}919$, гипотеза экспоненциальности не отклоняется.

3.3.10. Критерий Климко–Антла–Радемакера–Рокетта

В [311] рассмотрен критерий экспоненциальности, основанный на проверке равенства единице коэффициента формы β в распределении Вейбулла с функцией

$$F(x) = 1 - \exp\left\{-\left(\frac{x-\mu}{\alpha}\right)^\beta\right\}.$$

Известно (см. раздел 1.1.5), что при $\beta = 1$ распределение Вейбулла переходит в экспоненциальное распределение. Поэтому проверка гипотезы $H_0\colon \beta = 1$ против альтернативы $H_1: \beta > 1$ эквивалентна проверке гипотезы экспоненциальности распределения против альтернативы, утверждающей, что распределение является вейбулловским.

Для оценки параметра β может быть использована аппроксимация [312] $\tilde{c} \approx \nu^{-1,075}$, где $\nu = \frac{s}{\bar{x}}$ — коэффициент вариации. В общем виде для трехпараметрического распределения Вейбулла имеем

$$\nu = \frac{\left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{1}{2}}}{\frac{1}{n}\sum_{i=1}^{n}\left\{x_i - \left[\min_i x_i - \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(x_i - \min_i x_i\right)\right]\right\}}.$$

Критические значения статистики $c = \sqrt{n}\,(\tilde{c} - 1)$ приведены в табл. 89. Если $c > c_\alpha(n)$, то гипотеза $\beta = 1$ (гипотеза экспоненциальности) отклоняется.

Таблица 89

Критические значения статистики
$c = \sqrt{n}\,(\tilde{c} - 1)$
(α — уровень значимости) [311]

n	α			
	0,20	0,10	0,05	0,02
10	1,14	1,59	2,05	2,68
20	1,00	1,41	1,80	2,26
30	0,94	1,34	1,68	2,09
40	0,90	1,28	1,60	1,97
50	0,86	1,24	1,58	1,96
100	0,81	1,19	1,51	1,87

Задача 135. *В условиях задачи* 125 *проверить гипотезу экспоненциальности распределения вероятностей случайных величин критерием Климко–Антла–Радемакера–Рокетта на уровне значимости* $\alpha = 0{,}05$.

Имеем

$$s = \left[\frac{1}{9}\cdot\sum_{i=1}^{10}(x_i - \bar{x})^2\right]^{\frac{1}{2}} = 11{,}8157; \quad \hat{\mu} = 1 - \frac{1}{10\cdot 9}\cdot(0+1+3+\ldots+28+34) = -0{,}3889;$$

$$\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - \hat{\mu}) = 13{,}1111; \quad \tilde{c} = \left(\frac{11{,}8157}{13{,}1111}\right)^{-1{,}075} = 1{,}1183; \quad c = 10\cdot(\tilde{c} - 1) = 0{,}3741.$$

Из табл. 89 для $n = 10$ и $\alpha = 0{,}05$ находим $c_{0,95}(10) = 2{,}05$. Так как $c = 0{,}3741 < c_{0,95}(10) = 2{,}05$, гипотеза $\beta = 1$ (гипотеза экспоненциальности) не отклоняется.

3.3.11. Критерий Холлендера–Прошана

В задачах теории надежности экспоненциальное распределение наработки на отказ $f(x) = \lambda\exp(-\lambda x)$ характеризуется значением параметра $\lambda = \mathrm{const}$, т. е. постоянством интенсивности отказов изделия во времени. Отсюда следует, что вероятность безотказной работы изделия за время $\Delta t \to \exp(-\lambda\Delta t)$ определяется только промежутком времени Δt и не зависит от того, работало ли изделие раньше или нет. Другими словами, вероятность безотказной работы нового изделия и изделия, проработавшего часть времени, должна быть одинакова. Проверка этого обстоятельства и является целью критерия Холлендера–Прошана, рассмотренного в [313, 314].

Процедура вычисления статистики критерия Холлендера–Прошана включает в себя вычисление для ряда случайных величин $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ статистики

$$T = \sum_{i>j>k}\varphi(x_i, x_j + x_k), \quad \text{где} \quad \varphi(a,b) = \begin{cases} 1 & \text{при } a > b; \\ 0 & \text{при } a < b. \end{cases}$$

Суммирование проводится по всем $\dfrac{n(n-1)(n-2)}{6}$ тройкам i, j, k, для которых $i > j > k$.

На уровне значимости α гипотеза экспоненциальности принимается, если $t_1(\alpha) \leqslant T \leqslant t_2(\alpha)$, где $t_1(\alpha)$ и $t_2(\alpha)$ — граничные значения для заданных n и α, приведенные в табл. 90.

Таблица 90

Граничные значения $t_1(\alpha)$ и $t_2(\alpha)$ критерия Холлендера–Прошана [313]

n	Уровень значимости α							
	0,01		0,025		0,05		0,10	
	t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2
4							1	10
5	2		1	2			4	19
6	7		5	7		20	9	33
7	15	35	11	14		34	18	52
8	27	55	21	54	25	53	30	76
9	42	81	34	80	40	78	47	107
10	63	114	52	112	60	110	69	146
11	89	155	176	152	86	150	97	193
12	122	205	405	201	117	198	131	249
13	162	264	141	260	157	255	174	315
14	209	334	185	328	204	322	223	392
15	260	415	236	408	259	401	282	480
16	330	508	298	499	323	491	350	580
17	405	613	368	603	397	593	429	693
18	490	732	446	720	480	709	518	820
19	594	866	538	852	577	838	619	961
20	1250	1015	642	998	685	982	732	1918
25	2320	2018	1351	1986	1427	1956	1507	3349
30	3850	3521	2463	3461	2574	3411	2704	5359
35	5947	5625	4064	5546	4215	5464	4394	8049
40	8665	8418	6214	8301	6434	8189	6700	11502
45	12170	12012	9040	11864	9341	11709	9686	15823
50		16519	12661	16310	13020	16085	13439	

Для случаев совпадения $x_i = x_j + x_k$ используется замена $\varphi(x_i, x_j + x_k)$ на $\varphi^*(x_i, x_j + x_k)$, где

$$\varphi^*(a,b) = \begin{cases} 1 & \text{при } a > b; \\ \dfrac{1}{2} & \text{при } a = b; \\ 0 & \text{при } a < b. \end{cases}$$

Для $n > 20$ может быть использована аппроксимация

$$T^* = \frac{T - \mathbf{M}(T)}{\sqrt{\mathbf{D}(T)}},$$

где

$$\mathbf{M}(T) = \frac{n(n-1)(n-2)}{8};$$

$$\mathbf{D}(T) = \frac{3n(n-1)(n-2)}{2}\left[\frac{5(n-3)(n-4)}{2592} + \frac{7(n-3)}{432} + \frac{1}{48}\right].$$

Случайная величина T^* распределена при $n \geqslant 20$ как стандартная нормальная случайная величина, и гипотеза экспоненциальности в этом случае принимается на уровне значимости α, если $|T^*| < u_{1-\frac{\alpha}{2}}$, где u_γ — γ-квантиль стандартного нормального распределения.

Задача 136. *Для данных задачи* 125 *проверить гипотезу экспоненциальности критерием Холлендера–Прошана на уровне значимости* $\alpha = 0{,}05$.

Результаты вычислений представлены в таблице, из которой находим

$$T = \sum_{i>j>k} \varphi(x_i, x_j + x_k) = 102.$$

Из табл. 90 имеем $t_1(0{,}05) = 60$ и $t_2(0{,}05) = 110$.

Так как $t_1 = 60 < T = 102 < t_2 = 110$, гипотеза экспоненциальности не отклоняется.

i	j	k	x_i	$x_j + x_k$	$\varphi(\varphi^*)$	i	j	k	x_i	$x_j + x_k$	$\varphi(\varphi^*)$
3	2	1	4	3	1	5	4	2	9	7	1/2
4	2	1	5	3	1	6	4	2	11	7	1
5	2	1	9	3	1	7	4	2	18	7	1
6	2	1	11	3	1	8	4	2	21	7	1
7	2	1	18	3	1	9	4	2	29	7	1
8	2	1	21	3	1	10	4	2	35	7	1
9	2	1	29	3	1	6	5	2	11	11	1/2
10	2	1	35	3	1	7	5	2	18	11	1
4	3	1	5	5	1/2	8	5	2	21	11	1
5	3	1	9	5	1	9	5	2	29	11	1
6	3	1	11	5	1	10	5	2	35	11	1
7	3	1	18	5	1	7	6	2	18	13	1
8	3	1	21	5	1	8	6	2	21	13	1
9	3	1	29	5	1	9	6	2	29	13	1
10	3	1	35	5	1	10	6	2	35	13	1
5	4	1	9	6	1	8	7	2	21	20	1
6	4	1	11	6	1	9	7	2	29	20	1
7	4	1	18	6	1	10	7	2	35	20	1
8	4	1	21	6	1	9	8	2	29	23	1
9	4	1	29	6	1	10	8	2	35	23	1
10	4	1	35	6	1	10	9	2	35	31	1
6	5	1	11	10	1	5	4	3	9	9	1/2
7	5	1	18	10	1	6	4	3	11	9	1
8	5	1	21	10	1	7	4	3	18	9	1
9	5	1	29	10	1	8	4	3	21	9	1
10	5	1	35	10	1	9	4	3	29	9	1
7	6	1	18	12	1	10	4	3	35	9	1
8	6	1	21	12	1	6	5	3	11	13	0
9	6	1	29	12	1	7	5	3	18	13	1
10	6	1	35	12	1	8	5	3	21	13	1
8	7	1	21	19	1	9	5	3	29	13	1
9	7	1	29	19	1	10	5	3	35	13	1
10	7	1	35	19	1	7	6	3	18	15	1
9	8	1	29	22	1	8	6	3	21	15	1
10	8	1	35	22	1	9	6	3	29	15	1
10	9	1	35	30	1	10	6	3	25	15	1
4	3	2	5	6	0	8	7	3	21	22	0
5	3	2	9	6	1	9	7	3	29	22	1
6	3	2	11	6	1	10	7	3	35	22	1
7	3	2	18	6	1	9	8	3	29	25	1
8	3	2	21	6	1	10	8	3	35	25	1
9	3	2	29	6	1	10	9	3	35	33	1
10	3	2	35	6	1	6	5	4	11	14	0

Окончание

i	j	k	x_i	$x_i + x_k$	$\varphi(\varphi^*)$	i	j	k	x_i	$x_i + x_k$	$\varphi(\varphi^*)$
7	5	4	18	14	1	10	6	5	35	20	1
8	5	4	21	14	1	8	7	5	21	27	0
9	5	4	29	14	1	9	7	5	29	27	1
10	5	4	35	14	1	10	7	5	35	27	1
7	6	4	18	16	1	9	8	5	29	30	0
8	6	4	21	16	1	10	8	5	35	30	1
9	6	4	29	16	1	10	9	5	35	34	1
10	6	4	35	16	1	8	7	6	21	29	0
8	7	4	21	23	0	9	7	6	29	29	$1x2$
9	7	4	29	23	1	10	7	6	35	29	1
10	7	4	35	23	1	9	8	6	29	32	0
9	8	4	29	26	1	10	8	6	35	32	1
10	8	4	35	26	1	10	9	6	35	40	0
10	9	4	35	34	1	9	8	7	29	39	0
7	6	5	18	20	0	10	8	7	35	39	0
8	6	5	21	20	1	10	9	7	35	47	0
9	6	5	29	20	1	10	9	8	35	50	0

Для нормального приближения имеем

$$\mathbf{M}(T) = \frac{10 \cdot 9 \cdot 8}{8} = 90;$$

$$\mathbf{D}(T) = \frac{3 \cdot 10 \cdot 9 \cdot 8}{2} \cdot \left(\frac{5 \cdot 6 \cdot 7}{2592} + \frac{7 \cdot 7}{432} + \frac{1}{48} \right) = 232{,}5 \ (\sqrt{\mathbf{D}(T)} = 15{,}248);$$

$$T^* = \frac{102 - 90}{15{,}248} = 0{,}787.$$

Из табл. 1 имеем $u_{1-\frac{0,05}{2}} = u_{0,975} = 1{,}96$.

Так как $|T^*| = 0{,}852 < 1{,}96$, гипотеза экспоненциальности не отклоняется.

3.3.12. Критерий Кочара

Критерий, предложенный Кочаром [315], применяется для проверки постоянства интенсивности отказов против альтернативы монотонного увеличения интенсивности отказов (т. е. „старения“, износа изделия). Статистика критерия (имеется в виду упорядоченный по возрастанию ряд $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$) имеет вид

$$T_n = \sqrt{\frac{108n}{17}} \, \frac{\sum\limits_{i=1}^{n} J\left(\dfrac{i}{n+1}\right)}{\sum\limits_{i=1}^{n} x_i},$$

где $J\left(\dfrac{i}{n+1}\right) = 2\left(\dfrac{n+1-i}{n+1}\right)\left[1 - \ln\left(\dfrac{n+1-i}{n+1}\right)\right] - 1$.

При $n \geqslant 20$ распределение T_n удовлетворительно аппроксимируется стандартным нормальным распределением. Если $T_n \geqslant T_n(\alpha)$, то гипотеза экспоненциальности отклоняется ($T_n(\alpha)$ — критические значения, приведенные для различных уровней значимости α в табл. 91).

Критерий Кочара обладает высокой мощностью против альтернатив износа среди критериев подобного типа, рассмотренных в [316–321].

Таблица 91

Критические значения $T_n(\alpha)$ критерия Кочара (α — уровень значимости) [315]

n	α		n	α	
	0,01	0,05		0,01	0,05
2	2,2604	2,2265	12	2,8678	2,4724
3	2,5321	2,3923	13	2,8698	2,4627
4	2,6628	2,4536	14	2,8705	2,4528
5	2,7347	2,4863	15	2,8703	2,4429
6	2,7814	2,5006	16	2,8692	2,4332
7	2,8128	2,5046	17	2,8676	2,4235
8	2,8338	2,5029	18	2,8566	2,4141
9	2,8481	2,4976	19	2,8631	2,4049
10	2,8577	2,4903	20	2,8604	2,3559
11	2,8640	2,4817			

Задача 137. *Проверить экспоненциальность распределения для данных задачи* 125 *критерием Кочара на уровне значимости* $\alpha = 0{,}05$.

Имеем

$$J\left(\frac{1}{n+1}\right) = 2 \cdot \frac{10}{11} \cdot \left(1 - \ln\frac{10}{11}\right) - 1 = 0{,}99147;$$

$$J\left(\frac{2}{n+1}\right) = 2 \cdot \frac{9}{11} \cdot \left(1 - \ln\frac{9}{11}\right) - 1 = 0{,}96473;$$

$$J\left(\frac{3}{11}\right) = 0{,}91775; \quad J\left(\frac{4}{11}\right) = 0{,}84798; \quad J\left(\frac{5}{11}\right) = 0{,}75215; \quad J\left(\frac{6}{11}\right) = 0{,}62588;$$

$$J\left(\frac{7}{11}\right) = 0{,}46298; \quad J\left(\frac{8}{11}\right) = 0{,}25415; \quad J\left(\frac{9}{11}\right) = -0{,}01645; \quad J\left(\frac{10}{11}\right) = -0{,}38220;$$

$$\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \cdot x_i = 0{,}99147 \cdot 1 + 0{,}96473 \cdot 2 + \ldots$$
$$\ldots + 0{,}25415 \cdot 21 - 0{,}01645 \cdot 29 - 0{,}38220 \cdot 25 = 24{,}31227.$$

Получаем

$$T_n = \left(\frac{108 \cdot 10}{17}\right)^{\frac{1}{2}} \cdot \frac{24{,}31227}{135} = 1{,}435.$$

Из табл. 91 для $n = 10$ и $\alpha = 0{,}95$ находим $T_n(0{,}05) = 2{,}4903$.

Так как $T_n = 1{,}435 < T_n(0{,}05) = 2{,}4903$, гипотеза экспоненциальности распределения вероятностей случайных величин не отклоняется.

3.3.13. Критерий Эппса–Палли–Чёрго–Уэлча

Как и критерий Кочара (см. раздел 3.3.12), критерий Эппса–Палли [322] предполагает проверку гипотезы экспоненциальности против альтернативы монотонного изменения интенсивности отказов.

Критерий, подобно критерию Муроты–Такеучи для проверки нормальности распределения (см. раздел 3.2.2.17), использует выборочную характеристическую функцию

$$c_n(t) = \frac{1}{n} \sum_j \exp(itx_j)\,.$$

Статистика критерия имеет вид

$$c = \sqrt{48n}\left(\frac{1}{n}\sum_{i=1}^{n}\exp\left(-\frac{x_i}{\bar{x}}\right) - \frac{1}{2}\right).$$

Гипотеза экспоненциальности не отклоняется, если $c_1(\alpha) < c < c_2(\alpha)$, где $c_1(\alpha)$ и $c_2(\alpha)$ — критические значения, соответствующие доверительной вероятности α; они приведены в табл. 92.

Таблица 92

Критические значения $c_1(\alpha)$ и $c_2(\alpha)$ критерия Эппса–Палли (α — доверительная вероятность)

n	α					
	0,95		0,975		0,99	
	c_1	c_2	c_1	c_2	c_1	c_2
7	−1,64	1,31	−1,81	1,71	−2,00	2,18
8	−1,65	1,35	−1,83	1,74	−2,02	2,22
9	−1,65	1,38	−1,84	1,77	−2,04	2,25
10	−1,66	1,40	−1,85	1,79	−2,06	2,26
11	−1,66	1,42	−1,86	1,81	−2,08	2,28
12	−1,66	1,44	−1,87	1,82	−2,09	2,29
14	−1,67	1,46	−1,88	1,84	−2,11	2,31
16	−1,67	1,48	−1,89	1,86	−2,13	2,32
18	−1,67	1,49	−1,90	1,87	−2,15	2,33
20	−1,67	1,50	−1,91	1,88	−2,16	2,33
22	−1,67	1,51	−1,91	1,89	−2,17	2,34
24	−1,68	1,52	−1,92	1,89	−2,18	2,34
26	−1,68	1,53	−1,92	1,89	−2,19	2,34
28	−1,68	1,53	−1,92	1,90	−2,19	2,34
30	−1,68	1,54	−1,93	1,91	−2,20	2,35
35	−1,68	1,55	−1,93	1,91	−2,21	2,35
40	−1,68	1,56	−1,94	1,92	−2,22	2,35
45	−1,68	1,57	−1,94	1,92	−2,23	2,35
50	−1,68	1,57	−1,94	1,93	−2,24	2,35
60	−1,68	1,58	−1,95	1,93	−2,25	2,35
80	−1,67	1,59	−1,95	1,94	−2,25	2,35
100	−1,67	1,60	−1,95	1,94	−2,27	2,35
120	−1,67	1,60	−1,95	1,94	−2,28	2,35
160	−1,67	1,60	−1,96	1,96	−2,29	2,35
200	−1,67	1,62	−1,96	1,96	−2,30	2,35

При $n > 200c \to N(0,1)$ критические значения $c_1(\alpha) = -U_\alpha$ и $c_2(\alpha) = U_\alpha$, где U_α — α-квантиль стандартного нормального распределения.

При сравнительной простоте вычисления c-критерий достаточно мощен, не уступая критериям Шапиро–Уилка и Кочара против большинства альтернатив. Критерий, аналогичный критерию Эппса–Палли, рассмотрен Чёрго и Уэлчем [323]. Его статистика имеет вид

$$M = \sup_{-0,45\leqslant t\leqslant 0,45}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\exp\left(\frac{tx_i}{s} - \frac{1}{1-t}\right)\right\}\right|, \quad \text{где} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Гипотеза экспоненциальности отклоняется, если $M \geqslant M_n(\alpha)$, где $M_n(\alpha)$ — критические значения, приведенные для различных уровней достоверности α в табл. 93.

Таблица 93

Критические значения $M_n(\alpha)$ критерия Чёрго–Уэлча

(α — доверительная вероятность)

n	α		
	0,90	0,95	0,00
20	1,11	1,45	2,02
50	1,05	1,23	1,67
100	1,00	1,16	1,45

Критерий по мощности сравним с критерием Эппса–Палли и равносилен ему, однако требует значительно бóльших вычислений.

Задача 138. *В условиях задачи 125 проверить гипотезу экспоненциальности распределения вероятностей случайных величин критерием Эппса–Палли при доверительной вероятности $\alpha = 0{,}95$.*

Имеем $\bar{x} = 13{,}5$ и вычисляем статистику критерия Эппса–Палли

$$c = \sqrt{48 \cdot 10} \times$$
$$\times \left\{ \frac{1}{10} \cdot \left[\exp\left(-\frac{1}{13{,}5}\right) + \exp\left(-\frac{2}{13{,}5}\right) + \exp\left(-\frac{4}{13{,}5}\right) + \ldots + \exp\left(-\frac{35}{13{,}5}\right) \right] - \frac{1}{2} \right\} = 0{,}3345.$$

Из табл. 92 находим $c_1(0{,}95) = -1{,}66$ и $c_2(0{,}95) = 1{,}38$.

Так как $c_1(0{,}95) = -1{,}66 < c = -0{,}3345 < c_2(0{,}95) = 1{,}38$, гипотеза экспоненциальности не отклоняется.

3.3.14. Критерий Бергмана

Этот критерий базируется на проверке постоянства интенсивности отказов против альтернативы „бутылкообразной“ зависимости интенсивности отказов от времени (с участком приработки, когда $\lambda'(t) < 0$, и износа, когда $\lambda'(t) > 0$). Критерий развит в [323–325].

Определим для выборки объема n: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ переменную

$$U_i = \frac{\sum\limits_{j=1}^{i} (n-j+1)(x_j - x_{j-1})}{\sum\limits_{j=1}^{n} (n-j+1)(x_j - x_{j-1})}, \qquad x_0 = 0.$$

Для экспоненциального распределения $\dfrac{i}{n}$ и U_i должны быть связаны строгой линейной зависимостью. Отклонения от такой зависимости указывают на отклонение от экспоненциальности исследуемого распределения и являются источником построения критерия для установления этого факта.

Бергман [324] предложил простой критерий проверки экспоненциальности распределения, основанный на числе пересечений прямой $U_i = \dfrac{i}{n}$ графиком зависимости $U_i = f\left(\dfrac{i}{n}\right)$.

Обозначим через K_n число пересечений „снизу“, когда для $i = 1, \ldots, n-1$ имеют место неравенства $U_i < \frac{i}{n}$ и $U_{i+1} \geqslant \frac{i+1}{n}$. Статистикой критерия является общее число пересечений, определямое как

$$\delta_1 = \begin{cases} 1, & \text{если } U_1 < \frac{1}{n}, \\ 0 & \text{в ином случае;} \end{cases} \qquad \delta_2 = \begin{cases} 1, & \text{если } U_{n-1} < \frac{n-1}{n}, \\ 0 & \text{в ином случае.} \end{cases}$$

Значения вероятности того, что $L_n \geqslant k$ при справедливости гипотезы экспоненциальности распределения, приведены в табл. 94.

Таблица 94

Значения $\mathbf{P}(L_n \geqslant k)$ [324]

n	k								
	1	2	3	4	5	6	7	8	9
10	0,8000	0,4913	0,2564	0,0959	0,0294	0,0058	0,0008	0,0000	0,0000
20	0,9000	0,7143	0,5324	0,3416	0,2083	0,1068	0,0524	0,0211	0,0081
30	0,9333	0,8019	0,6628	0,4964	0,3617	0,2353	0,1504	0,0846	0,0471
40	0,9500	0,8485	0,7369	0,5945	0,4709	0,3427	0,2470	0,1618	0,1058
50	0,9600	0,8773	0,7844	0,6612	0,5499	0,4272	0,3299	0,2360	0,1690
60	0,9667	0,8969	0,8174	0,7093	0,6089	0,4938	0,3989	0,3023	0,2295
70	0,9714	0,9112	0,8417	0,7455	0,6545	0,5473	0,4563	0,3600	0,2849
90	0,9778	0,9304	0,8750	0,7964	0,7202	0,6269	0,5449	0,4535	0,3787
100	0,9800	0,9372	0,8869	0,8149	0,7445	0,6573	0,5795	0,4914	0,4180

Задача 139. *Проверить гипотезу экспоненциальности в условиях задачи* 125 *критерием Бергмана на уровне значимости* $\alpha = 0{,}05$.

Имеем $\sum_{j=1}^{10}(n-j+1)(x_j - x_{j-1}) = 135$.

Результаты вычислений сведем в таблицу:

i	$\frac{i}{n}$	x_i	U_i	i	$\frac{i}{n}$	x_i	U_i
1	0,1	1	0,0740	6	0,6	11	0,5630
2	0,2	2	0,1417	7	0,7	18	0,7703
3	0,3	4	0,2592	8	0,8	21	0,8370
4	0,4	5	0,3111	9	0,9	29	0,9555
5	0,5	9	0,4889	10	1,0	35	1,0000

Из таблицы видно, что

$$\delta_1 = 1 \ \left(U_1 = 0{,}074 < \frac{1}{10} = 0{,}1\right), \quad \delta_2 = 0 \ \left(U_9 = 0{,}9555 > \frac{9}{10} = 0{,}9\right),$$

$$K_n = 1 \ \left(U_6 = 0{,}5630 < \frac{6}{10} = 0{,}6 \text{ и } U_7 = 0{,}7703 > \frac{7}{10} = 0{,}7\right).$$

Имеем $L_n = 1 - 1 + 2 \cdot (1 - 0) + 0 - 1 = 1$. Из табл. 94 находим $\mathbf{P}(L_n \geqslant 1) = 0{,}80$, т. е. при справедливости гипотезы экспоненциальности эта вероятность высока и намного превосходит заданный уровень значимости. Следовательно, гипотеза экспоненциальности не отклоняется.

3.3.15. Критерий Шермана

Статистика критерия имеет вид [327]

$$\omega_n = \frac{1}{2n}\frac{\sum\limits_{i=1}^{n}|x_i - \bar{x}|}{\bar{x}}.$$

Критические значения статистики Шермана приведены в табл. 95.

Таблица 95

Критические значения $\omega_n(\alpha)$ статистики Шермана
(α — доверительная вероятность) [328]

n	α			n	α		
	0,90	0,95	0,99		0,90	0,95	0,99
1	0,450	0,475	0,495	11	0,442	0,469	0,521
2	0,484	0,537	0,609	12	0,440	0,466	0,516
3	0,467	0,518	0,614	13	0,438	0,463	0,511
4	0,468	0,509	0,589	14	0,436	0,460	0,506
5	0,462	0,502	0,574	15	0,434	0,458	0,502
6	0,458	0,494	0,562	16	0,433	0,455	0,498
7	0,454	0,488	0,551	17	0,431	0,453	0,495
8	0,451	0,482	0,542	18	0,430	0,451	0,491
9	0,448	0,477	0,534	19	0,429	0,449	0,489
10	0,445	0,473	0,527	20	0,427	0,448	0,486

При $n > 20$ критерий Шермана удовлетворительно аппроксимируется нормальным распределением со средним $\mathbf{M}(\omega_n)$ и дисперсией $\mathbf{D}(\omega_n)$, где

$$\mathbf{M}(\omega_n) = \left(\frac{n}{n+1}\right)^{n+1} \approx \frac{1}{e} = 0{,}36788;$$

$$\mathbf{D}(\omega_n) = \frac{2n^{n+2} + n(n-1)^{n+2}}{(n+2)(n+1)^{n+2}} - \left(\frac{n}{n+1}\right)^{2n+2} \approx \frac{2e-5}{e^2}\frac{1}{n} \approx \frac{0{,}05908}{n}.$$

Следовательно, случайная величина $\omega_n^* = \dfrac{\omega_n - \mathbf{M}(\omega_n)}{\sqrt{\mathbf{D}(\omega_n)}}$ имеет стандартное нормальное распределение. В [329] рассмотрена аппроксимация распределения ω_n-статистики F-распределением Фишера. Там же предложена очень эффективная нормальная аппроксимация

$$\tilde{\omega}_n = U - \frac{0{,}0955}{\sqrt{n}}\left(U^2 - 1\right), \quad \text{где} \quad U = \frac{\omega_n - 0{,}3679\left(1 - \dfrac{1}{2n}\right)}{0{,}2431\sqrt{n}\left(1 - \dfrac{0{,}605}{n}\right)}.$$

Статистика $\tilde{\omega}_n$ хорошо аппроксимируется стандартным нормальным распределением уже при $n > 20$.

Если $\omega_n > \omega_n(\alpha)$ или $\tilde{\omega}_n > u_\alpha$, то с доверительной вероятностью α гипотеза экспоненциальности отклоняется (u_α — α-квантиль стандартного нормального распределения).

Задача 140. *Проверить гипотезу экспоненциальности для данных задачи 125 критерием Шермана при доверительной вероятности $\alpha = 0{,}95$.*

Имеем

$$\bar{x} = 13{,}5; \quad \omega_n = \frac{1}{2 \cdot 10 \cdot 13{,}5} \cdot (|1 - 13{,}5| + |2 - 13{,}5| + \ldots + |35 - 13{,}5|) = 0{,}4.$$

Из табл. 95 для $n = 10$ находим $\omega_n(0{,}95) = 0{,}473$.

Так как $\omega_n = 0{,}400 < \omega_n(0{,}95) = 0{,}473$, гипотеза экспоненциальности не отклоняется.

Используем теперь нормальную аппроксимацию

$$U = \frac{0{,}4 - 0{,}3679 \cdot \left(1 - \frac{1}{2 \cdot 10}\right)}{0{,}2431 \cdot \sqrt{10} \cdot \left(1 - \frac{0{,}605}{10}\right)} = 0{,}069914;$$

$$\tilde{\omega}_n = 0{,}069914 - \frac{0{,}0995}{\sqrt{10}} \cdot \left(0{,}069914^2 - 1\right) = 0{,}1012,$$

что существенно меньше, чем $u_{0{,}95} = 1{,}645$ (однако следует помнить, что пример демонстрационный и нормальным приближением следует пользоваться только при $n > 20$).

3.3.16. Критерий наибольшего интервала

Статистика критерия имеет вид [140]

$$\eta_n = \frac{\max\limits_{1 \leqslant i \leqslant n} (x_i - x_{i-1})}{\sum\limits_{i=1}^{n} x_i}.$$

Статистика η_n совпадает с известной статистикой Кохрана для проверки однородности нескольких дисперсий (при числе степеней свободы $f = 2$). Критические значения $\eta_n(\alpha)$ приведены в табл. 96.

Таблица 96

Критические значения $\eta_n(\alpha)$ статистики η_n [25]

n	Доверительная вероятность α		n	Доверительная вероятность α	
	0,95	0,99		0,95	0,99
2	0,975	0,995	12	0,392	0,475
3	0,871	0,942	15	0,335	0,407
4	0,768	0,864	20	0,270	0,330
5	0,684	0,788	24	0,235	0,287
6	0,616	0,722	30	0,198	0,241
7	0,561	0,664	40	0,158	0,191
8	0,561	0,664	60	0,113	0,137
9	0,477	0,573	120	0,063	0,076
10	0,445	0,536			

Задача 141. *Проверить гипотезу экспоненциальности критерием наибольшего интервала для данных задачи 125 при $\alpha = 0{,}95$.*

Имеем $\max(x_i - x_{i-1}) = 8$; $\sum x_i = 135$; $\eta_n = \dfrac{8}{135} = 0{,}0592$.

Из табл. 96 имеем $\eta_n(0{,}95) = 0{,}445$.

Так как $\eta_n = 0{,}0592 < \eta_n(0{,}95) = 0{,}445$, гипотеза экспоненциальности не отклоняется.

3.3.17. Критерий Хартли [330]

Это один из наиболее простых в вычислительном отношении критериев (однако его мощность, естественно, уступает рассмотренным ранее). Статистика критерия имеет вид

$$h(n) = \frac{\max\limits_{1 \leqslant i \leqslant n} x_i}{\min\limits_{1 \leqslant i \leqslant n} x_i}.$$

При справедливости нулевой гипотезы $h(n)$ имеет χ^2-распределение с $f = 2$ степенями свободы. Критические значения $H_\alpha(n)$ статистики Хартли приведены в табл. 97.

Таблица 97

Критические значения $h_\alpha(n)$ статистики Хартли

(α — доверительная вероятность) [140]

n	α		n	α		n	α	
	0,95	0,99		0,95	0,99		0,95	0,99
2	39	199	6	266	1362	10	550	2813
3	88	448	7	333	1705	11	626	3204
4	142	729	8	403	2063	12	704	3605
5	202	1036	9	475	2432			

При $h(n) > H_\alpha(n)$ гипотеза экспоненциальности отклоняется.

Задача 142. *В условиях задачи* 125 *проверить гипотезу экспоненциальности критерием Хартли при* $\alpha = 0{,}95$.

Имеем $h(n) = \dfrac{\max x_i}{\min x_i} = \dfrac{35}{1} = 35$.

Для $n = 10$ из табл. 97 имеем $H_{0,95}(10) = 550$, что существенно больше полученного значения $h(n) = 35$. Следовательно, гипотеза экспоненциальности не отклоняется.

3.3.18. Критерий показательных меток

Применительно к задачам испытаний на надежность постоянство интенсивности отказов указывает на экспоненциальный закон распределения вероятностей значений временны́х промежутков между соседними отказами. В [310] для проверки экспоненциальности рассмотрен критерий показательных меток. Пусть имеются моменты появления отказов изделия $t_1, t_2, \ldots, t_n$ и порождаемая ими последовательность интервалов между отказами

$$\Delta t_1 = t_1 - t_2; \quad \Delta t_2 = t_3 - t_2; \quad \ldots; \quad \Delta t_i = t_{i+1} - t_i; \quad \ldots; \quad \Delta t_{n-1} = t_n - t_{n-1}.$$

Каждому интервалу ставится в соответствие его номер и так называемая метка, равная математическому ожиданию r-й порядковой статистики из единичного экспоненциального распределения в выборке объема $n - s_{r,n-1}$ (здесь r — порядковый номер интервала Δt_r в общем, упорядоченном по возрастанию ряду $(n-1)$ интервалов)

$$s_{r,n} = \sum_{j=1}^{r} \frac{1}{n - j + 1}.$$

Статистика критерия имеет вид [310]

$$s = \sum_{i=1}^{n} s_{i,n}(z_i - \bar{z}).$$

Критерий применяется при $n > 50$, его критические значения в этом случае вычисляются, исходя из того, что величина s имеет асимптотически нормальное распределение со средним $\mathbf{M}(s) = 0$ и дисперсией

$$\mathbf{D}(s) = \sum (z_i - \bar{z})\, k_n, \quad \text{где} \quad k_n = 1 - \frac{1}{n-1}\left(\frac{1}{n} + \frac{1}{n+1} + \ldots + \frac{1}{2}\right).$$

Задача 143. *Предположим, что ряд данных задачи* 125

$$x_i\colon\ 1,\ 2,\ 4,\ 5,\ 9,\ 11,\ 18,\ 21,\ 29,\ 35$$

представляет собой моменты отказов изделий. Проверить гипотезу экспоненциальности распределения x_i критерием показательных меток при доверительной вероятности $\alpha = 0{,}95$ (при совпадении рангов интервалов r можно использовать их случайную последовательную нумерацию).

Результаты расчетов приведены в таблице:

i	x_i	Δx_i	r	$s_{r,n}$	z_i	$z_i - \bar{z}$	$(z_i - \bar{z})^2$	$s_{r,n}(z_i - \bar{z})$
1	1							
2	2	1	1	0,1000	1	−4	16	−0,4000
3	4	2	3	0,3361	2	−3	9	−1,0083
4	5	1	2	0,2111	3	−2	4	−0,4222
5	9	4	6	0,8456	4	−1	1	−0,8456
6	11	2	4	0,4790	5	0	0	0
7	18	7	8	1,4290	6	1	1	1,4290
8	21	3	5	0,6456	7	2	4	1,2912
9	29	8	9	1,9290	8	3	9	5,7870
10	35	6	7	1,0956	9	4	16	4,3824

Для $n = 9$ $\quad k_n = 1 - \frac{1}{8}\cdot\left(\frac{1}{9} + \frac{1}{8} + \ldots + \frac{1}{3} + \frac{1}{2}\right) = 0{,}7138.$

Далее $s = 10{,}2135$; $\mathbf{M}(s) = 0$; $\mathbf{D}(S) = 0{,}7138 \cdot 60 = 42{,}828$ $(\sqrt{\mathbf{D}(s)} = 6{,}544)$.
Случайная величина

$$s^* = \frac{s - \mathbf{M}(s)}{\sqrt{\mathbf{D}(s)}} = \frac{10{,}2135}{6{,}544} = 1{,}5607$$

имеет стандартное нормальное распределение, 95%-я квантиль которого равна $u_{0,95} = 1{,}645$ (см. табл. 1). Так как $s^* = 1{,}5607 < u_{0,95} = 1{,}645$, гипотеза экспоненциальности не отклоняется.

3.3.19. Ранговый критерий независимости интервалов

Характеристической особенностью экспоненциального распределения случайных величин является независимость интервалов между двумя соседними случайными величинами $\Delta x_i = x_i - x_{i-1}$.

Для выборок объема $n > 30$ в качестве оценки такой независимости может служить ранговый смешанный момент [296, 310]

$$R_1 = \sum_{i=1}^{n-1} r_{i+1} r_i,$$

где r_j — ранг j-го интервала в общем упорядоченном ряду значений интервалов.

Например, если были получены величины $x_1 = 3$, $x_2 = 1$, $x_3 = 5$, $x_4 = 2$, $x_5 = 11$, то после упорядочения по величине мы имеем $x_2 = 1$, $x_4 = 2$, $x_1 = 3$, $x_3 = 4$ и $x_5 = 5$.

При справедливости гипотезы экспоненциальности случайная величина

$$R_1^* = \frac{R - \mathbf{M}(R_1)}{\mathbf{D}(R_1)},$$

где

$$\mathbf{M}(R_1) = \frac{1}{12}(n-1)(n+1)(3n+2);$$

$$\mathbf{D}(R_1) = \frac{(n+1)\left(5n^6 + 21n^5 + 501n^4 - 823n^3 + 1102n^2 - 68n - 240\right)}{720(n-2)(n-3)},$$

имеет стандартное нормальное распределение.

Отсюда следует правило: если $|R_1^*| < u_{1-\frac{\alpha}{2}}$, где $u_{1-\frac{\alpha}{2}}$ — $\left(1 - \dfrac{\alpha}{2}\right)$-квантиль стандартного нормального распределения, то на уровне значимости α гипотеза экспоненциальности не отклоняется.

В [310] рассмотрена модификация этого критерия в форме

$$\tilde{R}_1 = \sum_{i=1}^{n-1} s_{r_{i+1},n} s_{r_i,n},$$

где $s_{r,n} = \sum\limits_{i=1}^{r} \dfrac{1}{n-i+1}$ — математическое ожидание r-й порядковой статистики из единичного нормированного экспоненциального распределения.

По аналогии рассматривается статистика

$$\tilde{R}_1^* = \frac{\tilde{R}_1 - \mathbf{M}(\tilde{R}_1)}{\sqrt{\mathbf{D}(\tilde{R}_1)}},$$

где

$$\mathbf{M}(\tilde{R}_1) = n - 2 + \frac{\ln n + ,05772}{n} + \frac{1}{2n^2}; \quad \mathbf{D}(\tilde{R}_1) = \frac{n^3 - 6n^2 + 24n}{(n-2)(n-3)} - 2\ln n.$$

Правило проверки гипотезы экспоненциальности аналогично случаю критерия R_1^*. При совпадении рангов $r_i = r_j$ используются средние ранги (например, в случае $r_i = r_j$ для обеих величин принимается ранг $\dfrac{r_i + r_j}{2}$). Следует помнить, что эти критерии обладают приемлемой мощностью только при $n \geqslant 30$.

Задача 144. *Проверить гипотезу экспоненциальности распределения вероятностей случайных величин для данных задачи* 125 *критериями независимости интервалов* ($\alpha = 0{,}95$).

Для ряда

$$x_i \ (n = 10){:}\ 1,\ 2,\ 4,\ 5,\ 9,\ 11,\ 18,\ 21,\ 29,\ 35$$

имеем ряд интервалов

$$\Delta x_i \ (n = 9){:}\ 1,\ 2,\ 1,\ 4,\ 2,\ 7,\ 3,\ 8,\ 6$$

и соответствующий им ряд рангов

$$r_i{:}\ 1{,}5;\ 3{,}5;\ 1{,}5;\ 6;\ 3{,}5;\ 8;\ 5;\ 9;\ 7.$$

Вычисляем

$$R_1 = \sum_{i=1}^{n-1} r_{i+1} \cdot r_i = 3{,}5 \cdot 1{,}5 + 1{,}5 \cdot 3.5 + \ldots + 9 \cdot 5 + 7 \cdot 9 = 216{,}5;$$

$$\mathbf{M}(R_1) = \frac{1}{720 \cdot 7 \cdot 6} \times \times \left[10 \cdot \left(5 \cdot 9^6 + 21 \cdot 9^5 + 501 \cdot 9^4 - 823 \cdot 9^3 + 1102 \cdot 9^2 - 68 \cdot 9 - 240\right)\right] = 2206{,}5932.$$

Тогда $R_1^* = \dfrac{216{,}5 - 193{,}3}{46{,}9743} = 0{,}494$, что меньше $u_{1-\frac{\alpha}{2}} = u_{0{,}975} = 1{,}96$.

Следовательно, гипотеза экспоненциальности не отклоняется.

Вычислим теперь критерий $\tilde{R}_1$. Имеем

$$s_{r_1,n} = \frac{1}{9} = 0{,}1111; \quad s_{r_2,n} = 0{,}2361; \quad s_{r_3,n} = 0{,}3790; \quad s_{r_4,n} = 0{,}5456; \quad s_{r_5,n} = 0{,}7456;$$

$$s_{r_6,n} = 0{,}9956; \quad s_{r_7,n} = 1{,}329; \quad s_{r_8,n} = 1{,}829; \quad s_{r_9,n} = 2{,}829;$$

$$s_{r_{1,5},n} = \frac{s_{r_1,n} + s_{r_2,n}}{2} = 0{,}1736; \quad s_{r_{3,5},n} = \frac{s_{r_3,n} + s_{r_4,n}}{2} = 0{,}4626;$$

$$\tilde{R}_1 = 0{,}4626 \cdot 0{,}1736 + 0{,}1736 \cdot 0{,}4626 + 0{,}9956 \cdot 0{,}1736 + \ldots + 1{,}329 \cdot 2{,}829 = 8{,}8728;$$

$$\mathbf{M}(\tilde{R}_1) = 9 - 2 + \frac{1}{9} \cdot (\ln 9 + 0{,}5772) + \frac{1}{2 \cdot 9^2} = 7{,}3144;$$

$$\mathbf{D}(\tilde{R}_1) = \frac{1}{7 \cdot 6} \cdot (9^3 - 6 \cdot 9^2 + 24 \cdot 9) - 2\ln 9 = 6{,}534 \quad \left(\sqrt{\mathbf{D}(\tilde{R}_1)} = 2{,}556\right);$$

$$\tilde{R}_1^* = \frac{8{,}8728 - 7{,}3144}{2{,}556} = 0{,}6097.$$

Так как $\tilde{R}_1^* = 0{,}6097 < u_{1-\frac{\alpha}{2}} = u_{0{,}975} = 1{,}96$, гипотеза экспоненциальности распределения вероятностей случайных величин не отклоняется.

3.3.20. Критерии, основанные на трансформации экспоненциального распределения в равномерное

Если случайные величины $x_1, x_2, \ldots, x_i, \ldots, x_n$ распределены экспоненциально, то случайные величины $U_i = \dfrac{\sum\limits_{j=1}^{i} x_j}{\sum\limits_{j=1}^{n} x_j}$ должны быть распределены равномерно на интервале $[0, 1]$ [231]. В этом случае проверка равномерности распределения величин U_i эквивалентна проверке экспоненциальности распределения исходных величин x_i. Рассмотрим серию критериев такого типа.

3.3.20.1. Критерий $\bar{U}$

Простейший критерий, основан на статистике

$$\bar{U} = \frac{1}{n-1} \sum_{i=1}^{n-1} U_i.$$

Критические значения статистики $\bar{U}_\alpha$ приведены в табл. 98.

При $n \geqslant 15$ можно использовать тот факт, что величина

$$\bar{U}^* = \frac{1}{\sqrt{2(n-1)}}\left(\bar{U} - 0{,}5\right)$$

распределена как $N(0, 1)$ и ее квантили тогда могут быть вычислены через квантили стандартного нормального распределения.

Таблица 98

Критические значения $\bar{U}_\alpha$ статистики $\bar{U}$ (α — уровень значимости) [231]

n	α				n	α			
	0,01	0,025	0,05	0,10		0,01	0,025	0,05	0,10
5	0,176	0,221	0,262	0,312	19	0,343	0,367	0,388	0,412
6	0,208	0,250	0,287	0,332	21	0,351	0,374	0,394	0,417
7	0,232	0,271	0,306	0,347	26	0,366	0,387	0,405	0,426
8	0,251	0,288	0,320	0,359	31	0,387	0,397	0,413	0,432
9	0,266	0,301	0,332	0,368	36	0,378	0,404	0,420	0,437
10	0,279	0,312	0,341	0,376	41	0,394	0,411	0,425	0,441
11	0,290	0,322	0,350	0,382	46	0,400	0,416	0,429	0,445
13	0,308	0,337	0,363	0,393	51	0,405	0,420	0,433	0,448
15	0,322	0,349	0,373	0,401	61	0,420	0,432	0,443	0,456
17	0,330	0,359	0,381	0,407					

3.3.20.2. Критерий $\tilde{U}$

Наиболее простой в вычислительном отношении критерий, основанный на статистике

$$\tilde{U} = \frac{(n-r)\,U_r}{r(1-U_r)},$$

имеющей при справедливости гипотезы экспоненциальности F-распределение с $f_1 = 2r$ и $f_2 = 2(n-r)$ степенями свободы. Рекомендуется выбирать $r = \frac{n}{2}$ (n — четное) или $r = \frac{n+1}{2}$ (n — нечетное).

3.3.20.3. Критерий Гринвуда

Известен ряд весьма эффективных критериев, основанных на спейсингах величин U_i, равных $D_i = U_i - U_{i-1}$ ($i = 1, \ldots, n$; $U_0 - 0$; $U_n = 1$). Здесь U_i следует рассматривать как порядковые статистики (т. е. ряд U_i должен быть предварительно упорядочен по возрастанию).

Легко видеть, что

$$D_i = \frac{x_i}{\sum\limits_{i=1}^{n} x_i}.$$

В качестве примера критерия экспоненциальности, основанного на спейсингах D_i, приведем критерий Гринвуда [331]

$$G = n\sum_{i=1}^{n} D_i^2 = n\frac{\sum\limits_{i=1}^{n} x_i^2}{\left(\sum\limits_{i=1}^{n} x_i\right)^2}.$$

Гипотеза экспоненциальности не отклоняется, если $G_1(\alpha) \leqslant G \leqslant G_2(\alpha)$, где $G_1(\alpha)$ и $G_2(\alpha)$ — критические значения, приведенные в табл. 99.

Таблица 99

Критические значения $G_1(\alpha)$ и $G_2(\alpha)$ статистики Гринвуда

(α — доверительная вероятность) [332, 333]

n	$\alpha = 0{,}90$		$\alpha = 0{,}95$		$\alpha = 0{,}99$	
	G_1	G_2	G_1	G_2	G_1	G_2
3	0,694	1,539	0,680	1,673	0,672	1,780
4	0,825	1,852	0,796	2,075	0,776	2,269
5	0,923	2,037	0,895	2,311	0,855	2,560
6	0,997	2,160	0,954	2,461	0,919	2,737
7	1,055	2,246	1,009	2,559	0,973	2,849
8	1,104	2,306	1,060	2,615	1,017	2,921
9	1,145	2,349	1,095	2,670	1,055	2,967
10	1,180	2,381	1,129	2,700	1,088	2,997
11	1,211	2,404	1,159	2,717	1,117	3,008
13	1,272	2,441	1,234	2,693	1,198	3,015
15	1,312	2,457	1,272	2,691	1,233	3,014
17	1,346	2,464	1,304	2,691	1,263	3,003
19	1,375	2,466	1,332	2,685	1,288	2,988
21	1,400	2,465	1,356	2,677	1,311	2,970
26	1,451	2,456	1,405	2,651	1,358	2,920
31	1,490	2,443	1,444	2,624	1,395	2,873
41	1,548	2,415	1,502	2,573	1,453	2,790
51	1,589	2,389	1,544	2,531	1,495	2,723
61	1,621	2,367	1,577	2,495	1,529	2,669
81	1,666	2,331	1,625	2,441	1,579	2,587
101	1,698	2,304	1,659	2,400	1,616	2,528
201	1,781	2,226	1,750	2,289	1,714	2,371
501	1,858	2,147	1,836	2,183	1,811	2,228

Задача 145. *Проверить экспоненциальность распределения по данным задачи* 125 *критериями* $\tilde{U}$ *и Гринвуда на уровне значимости* $\alpha = 0{,}05$.

Имеем $\sum\limits_{j=1}^{n} x_j = 135$. Далее вычисляем

$$U_1 = \frac{x_1}{\sum x_i} = \frac{1}{135} = 0{,}00741; \quad U_2 = \frac{3}{135} = 0{,}02222; \quad U_3 = \frac{7}{135} = 0{,}05185;$$

$$U_4 = 0{,}08889; \quad U_5 = 0{,}15555; \quad U_6 = 0{,}23703; \quad U_7 = 0{,}37037; \quad U_8 = 0{,}5259;$$

$$U_9 = 0{,}74074; \quad U_{10} = 1{,}00 \quad \text{и} \quad \bar{U} = \frac{1}{9} \cdot \sum_{i=1}^{9} U_i = 0{,}2444.$$

Из табл. 98 для $n = 9$ и $\alpha = 0{,}05$ находим $\bar{U}_{0,05} = 0{,}332$.

Так как $\bar{U} = 0{,}2444 < \bar{U}_{0,05} = 0{,}332$, гипотеза экспоненциальности не отклоняется.

Применим критерий, для которого (при $r = \frac{10}{2} = 5$) имеем

$$\tilde{U} = \frac{(10-5) \cdot U_5}{5 \cdot (1 - U_5)} = \frac{5 \cdot 0{,}1555}{5 \cdot (1 - 0{,}1555)} = 0{,}1842.$$

При $\alpha = 0{,}05$, $f_2 = 2 \cdot (10 - 5) = 10$ имеем $F_{0,05}(10{,}10) = 2{,}978$.

Следовательно, $\tilde{U} = 0{,}18 < 2{,}98$ и гипотеза экспоненциальности не отклоняется.

Рассмотрим теперь критерий Гринвуда. Имеем

$$D_1 = U_1 = 0{,}00741; \quad D_2 = U_2 - U_1 = 0{,}01481; \quad D_3 = U_3 - U_2 = 0{,}0295;$$
$$D_4 = 0{,}03704; \quad D_5 = 0{,}0666; \quad D_6 = 0{,}08148; \quad D_7 = 0{,}13333; \quad D_8 = 0{,}15555;$$
$$D_9 = 0{,}08148; \quad D_{10} = 0{,}2539 \quad \text{и} \quad G = n \cdot \sum_{i=1}^{10} D_i^2 = 1{,}689.$$

Из табл. 99 имеем $G_1(0{,}95) = 1{,}129$ и $G_2(0{,}95) = 2{,}700$.

Так как $G_1(\alpha) = 1{,}29 < G = 1{,}689 < G_2(\alpha) = 2{,}700$, гипотеза экспоненциальности не отклоняется.

3.3.21. Критерий Манн–Фертига–Шуера для распределения Вейбулла

Распределение Вейбулла является обобщающим для экспоненциального распределения. Закон распределения Вейбулла

$$F(x) = 1 - \exp\left(-\frac{x}{\alpha}\right)^{\beta}$$

совпадает при $\beta = 1$ с экспоненциальным. Поэтому распределение Вейбулла часто рассматривается как альтернативное при проверке экспоненциальности распределения.

Учитывая изложенное, критерий согласия для распределения Вейбулла приводится здесь. Критерий был предложен авторами применительно к задаче испытаний изделий на долговечность.

Если $t_1, t_2, \ldots, t_r$ — первые r порядковых статистик наработки на отказ при испытаниях выборки изделий объема $n \geqslant r$, то статистика критерия имеет вид

$$K = \frac{\sum\limits_{i=\left[\frac{r}{2}+1\right]}^{r-1} a_i^{-1}(x_{i+1} - x_i)}{\sum\limits_{i=1}^{r} a_i^{-1}(x_{i+1} - x_i)},$$

где $x_i = \ln t_i$, $\left[\dfrac{r}{2}\right]$ — наибольшее целое число $\leqslant \dfrac{r}{2}$, a_i — коэффициенты, приведенные в табл. 100.

Гипотеза согласия эмпирического распределения с двухпараметрическим распределением Вейбулла отклоняется, если

$$K > K_\alpha(r, n)\,,$$

где $K > K_\alpha(r, n)$ — критическое значение статистики для доверительной вероятности α (при известных результатах отказов r изделий из n), приведенное в табл. 101.

Задача 146. *Проверить гипотезу согласия данных задачи* 125 *с двухпараметрическим распределением Вейбулла при доверительной вероятности* $\alpha = 0{,}95$.

Для ряда значений x_i имеем ряд

$$\ln x_i\colon\ 0;\ 0,693;\ 1,386;\ 1,609;\ 2,197;\ 2,398;\ 2,890;\ 3,044;\ 3,367;\ 3,555.$$

С учетом коэффициентов a_i из табл. 100 для $n = r = 10$ имеем

$$\sum_{i=6}^{9} \frac{\ln x_{i+1} - \ln x_i}{a_i} = \frac{2{,}890 - 2{,}398}{0{,}269493} + \ldots + \frac{3{,}555 - 3{,}367}{0{,}405316};$$

Таблица 100

Коэффициенты a_i K-критерия согласия Манн–Фертига–Шуера для распределения Вейбулла [95]

n	i	a_i	n	i	a_i	n	i	a_i
3	1	1,216395	11	6	0,251386	15	9	0,180266
	2	0,863046		7	0,243928		10	0,180072
4	1	1,150727		8	0,251548		11	0,180072
	2	0,706698		9	0,283879		12	0,186347
	3	0,679596		10	0,389071		13	0,239842
5	1	1,115718	12	1	1,044137		14	0,344309
	2	0,645384		2	0,547721	16	1	1,032617
	3	0,532445		3	0,385338		2	0,534521
	4	0,583273		4	0,307221		3	0,370021
6	1	1,093929		5	0,263737		4	0,289169
	2	0,612330		6	0,238797		5	0,242049
	3	0,474330		7	0,226264		6	0,212103
	4	0,442920		8	0,224477		7	0,192338
	5	0,522759		9	0,235630		8	0,179407
7	1	1,079055		10	0,269966		9	0,171667
	2	0,591158		11	0,375356		10	0,168476
	3	0,442789	13	1	1,040515		11	0,170026
	4	0,387289		2	0,543556		12	0,177619
	5	0,387714		3	0,380417		13	0,194859
	6	0,480648		4	0,301300		14	0,232350
8	1	1,068252		5	0,256437		15	0,336283
	2	0,577339		6	0,229515	17	1	1,030618
	3	0,422889		7	0,213966		2	0,532290
	4	0,356967		8	0,207205		3	0,367507
	5	0,334089		9	0,209131		4	0,286765
	6	0,349907		10	0,222667		5	0,238765
	7	0,449338		11	0,258323		6	0,208278
9	1	1,060046		12	0,363582		7	0,187813
	2	0,566942	14	1	1,037513		8	0,173951
	3	0,409157		2	0,540059		9	0,164928
	4	0,337763		3	0,376352		10	0,159891
	5	0,304777		4	0,296496		11	0,158624
	6	0,297949		5	0,250650		12	0,161559
	7	0,322189		6	0,222377		13	0,170132
	8	0,424958		7	0,204885		14	0,188005
10	1	1,053606		8	0,195165		15	0,225729
	2	0,559013		9	0,192209		16	0,329085
	3	0,399100		10	0,196679	18	1	1,028850
	4	0,324470		11	0,211875		2	0,530332
	5	0,286163		12	0,248409		3	0,365314
	6	0,269493		13	0,353334		4	0,283846
	7	0,271645	15	1	1,034894		5	0,235958
	8	0,300869		2	0,537085		6	0,205051
	9	0,405316		3	0,372934		7	0,184055
11	1	1,048411		4	0,292518		8	0,169504
	2	0,552769		5	0,255180		9	0,159564
	3	0,391410		6	0,216712		10	0,153263
	4	0,314705		7	0,197893		11	0,150176
	5	0,273245		8	0,186266		12	0,150333

Окончание таблицы 100

n	i	a_i	n	i	a_i	n	i	a_i
18	13	0,154313	21	10	0,140087	23	20	0,159966
	14	0,163630		11	0,134200		21	0,197679
	15	0,181971		12	0,130451		22	0,297435
	16	0,219825		13	0,128702	24	1	1,021431
	17	0,322580		14	0,129025		2	0,522233
19	1	1,027277		15	0,131756		3	0,356436
	2	0,528594		16	0,137659		4	0,274051
	3	0,363389		17	0,148341		5	0,225086
	4	0,281692		18	0,167481		6	0,192892
	5	0,233535		19	0,205352		7	0,170330
	6	0,202291		20	0,306285		8	0,153877
	7	0,180882	22	1	1,023439		9	0,141549
	8	0,165807		2	0,524405		10	0,132195
	9	0,155189		3	0,358790		11	0,125099
	10	0,147984		4	0,276618		12	0,119811
	11	0,143650		5	0,278950		13	0,116054
	12	0,142012		6	0,195983		14	0,113677
	13	0,143250		7	0,173760		15	0,112638
	14	0,148031		8	0,157692		16	0,113007
	15	0,157921		9	0,145834		17	0,114990
	16	0,176611		10	0,137052		18	0,119014
	17	0,214520		11	0,130662		19	0,125889
	18	0,316666		12	0,126260		20	0,137235
20	1	1,0,25866		13	0,123640		21	0,156679
	2	0,527046		14	0,122763		22	0,194285
	3	0,361682		15	0,123763		23	0,293473
	4	0,279798		16	0,127019	25	1	1,020551
	5	0,231417		17	0,133316		2	0,521285
	6	0,199905		18	0,144273		3	0,355415
	7	0,178167		19	0,163552		4	0,272945
	8	0,162684		20	0,201355		5	0,223885
	9	0,151549		21	0,301693		6	0,191578
	10	0,143674	23	1	1,022380		7	0,168899
	11	0,138448		2	0,523269		8	0,152286
	12	0,135580		3	0,357557		9	0,139789
	13	0,135306		4	0,275268		10	0,130219
	14	0,137120		5	0,226417		11	0,122871
	15	0,142527		6	0,194351		12	0,117274
	16	0,152861		7	0,171948		13	0,113132
	17	0,171810		8	0,155666		14	0,110268
	18	0,209721		9	0,143549		15	0,108598
	19	0,311257		10	0,134451		16	0,108124
21	1	1,024594		11	0,127667		17	0,108944
	2	0,525657		12	0,122768		18	0,111289
	3	0,360159		13	0,119503		19	0,115596
	4	0,278117		14	0,117764		20	0,122683
	5	0,229551		15	0,117577		21	0,134165
	6	0,197821		16	0,119120		22	0,153650
	7	0,175815		17	0,122799		23	0,191137
	8	0,160009		18	0,129416		24	0,289773
	9	0,148471		19	0,140590			

Таблица 101

Критические значения $K_\alpha(r,n)$ критерия Манн–Фертига–Шуера (α — доверительная вероятность) [95]

n	r	α 0,90	α 0,95	α 0,99	n	r	α 0,90	α 0,95	α 0,99	n	r	α 0,90	α 0,95	α 0,99	n	r	α 0,90	α 0,95	α 0,99
3	3	0,90	0,95	0,99	10	9	0,71	0,76	0,85	14	5	0,80	0,86	0,94	16	14	0,63	0,67	0,75
4	3	0,90	0,95	0,99		10	0,64	0,69	0,79		6	0,68	0,74	0,86		15	0,66	0,70	0,77
	4	0,67	0,76	0,89	11	3	0,90	0,95	0,99		7	0,75	0,81	0,89		16	0,62	0,66	0,73
5	3	0,90	0,95	0,99		4	0,68	0,77	0,90		8	0,66	0,73	0,82	17	3	0,90	0,95	0,99
	4	0,68	0,77	0,89		5	0,80	0,86	0,94		9	0,72	0,77	0,85		4	0,69	0,78	0,90
	5	0,79	0,96	0,94		6	0,68	0,75	0,86		10	0,65	0,70	0,79		5	0,80	0,87	0,94
6	3	0,90	0,95	0,99		7	0,75	0,81	0,89		11	0,69	0,74	0,82		6	0,68	0,74	0,85
	4	0,68	0,76	0,89		8	0,66	0,72	0,82		12	0,64	0,68	0,77		7	0,75	0,80	0,89
	5	0,80	0,86	0,94		9	0,71	0,77	0,85		13	0,67	0,72	0,79		8	0,66	0,72	0,81
	6	0,66	0,73	0,84		10	0,64	0,70	0,79		14	0,62	0,67	0,75		9	0,72	0,77	0,85
7	3	0,90	0,95	0,99		11	0,69	0,74	0,82	15	3	0,90	0,95	0,99		10	0,65	0,70	0,79
	4	0,68	0,77	0,89	12	3	0,90	0,95	0,99		4	0,69	0,78	0,90		11	0,69	0,74	0,82
	5	0,80	0,86	0,94		4	0,68	0,78	0,89		5	0,80	0,86	0,94		12	0,64	0,69	0,77
	6	0,67	0,74	0,85		5	0,80	0,86	0,94		6	0,67	0,75	0,86		13	0,67	0,72	0,80
	7	0,74	0,80	0,80		6	0,67	0,74	0,85		7	0,75	0,81	0,89		14	0,63	0,68	0,75
8	3	0,90	0,95	0,99		7	0,75	0,81	0,89		8	0,66	0,72	0,82		15	0,66	0,70	0,77
	4	0,68	0,77	0,90		8	0,66	0,72	0,82		9	0,72	0,77	0,85		16	0,62	0,.66	0,74
	5	0,80	0,88	0,94		9	0,71	0,77	0,85		10	0,65	0,70	0,79		17	0,65	0,69	0,75
	6	0,67	0,74	0,85		10	0,65	0,70	0,79		11	0,69	0,74	0,82	18	3	0,90	0,95	0,99
	7	0,74	0,80	0,89		11	0,69	0,74	0,82		12	0,64	0,68	0,77		4	0,68	0,77	0,90
	8	0,65	0,71	0,81		12	0,63	0,68	0,76		13	0,67	0,72	0,79		5	0,80	0,86	0,94
9	3	0,90	0,95	0,99	13	3	0,90	0,95	0,99		14	0,63	0,67	0,75		6	0,67	0,75	0,86
	4	0,68	0,77	0,89		4	0,68	0,77	0,89		15	0,66	0,70	0,77		7	0,75	0,81	0,89
	5	0,80	0,86	0,94		5	0,80	0,86	0,94	16	3	0,90	0,95	0,99		8	0,66	0,73	0,82
	6	0,67	0,75	0,86		6	0,68	0,75	0,86		4	0,69	0,78	0,89		9	0,72	0,77	0,85
	7	0,74	0,80	0,89		7	0,75	0,81	0,90		5	0,80	0,86	0,94		10	0,65	0,71	0,80
	8	0,66	0,72	0,82		8	0,66	0,72	0,82		6	0,68	0,75	0,86		11	0,69	0,74	0,82
	9	0,71	0,76	0,85		9	0,72	0,77	0,85		7	0,75	0,81	0,89		12	0,64	0,69	0,77
10	3	0,90	0,95	0,99		10	0,65	0,70	0,79		8	0,66	0,72	0,82		13	0,68	0,72	0,80
	4	0,68	0,77	0,90		11	0,69	0,74	0,82		9	0,72	0,77	0,85		14	0,63	0,67	0,76
	5	0,80	0,86	0,94		12	0,64	0,68	0,76		10	0,65	0,71	0,79		15	0,66	0,70	0,78
	6	0,68	0,75	0,85		13	0,67	0,72	0,79		11	0,69	0,74	0,82		16	0,62	0,66	0,74
	7	0,75	0,81	0,89	14	3	0,90	0,95	0,99		12	0,64	0,69	0,77		17	0,65	0,69	0,76
	8	0,66	0,72	0,81		4	0,68	0,77	0,90		13	0,68	0,72	0,80		18	0,62	0,65	0,71

Окончание таблицы 101

n	r	α 0,90	α 0,95	α 0,99	n	r	α 0,90	α 0,95	α 0,99	n	r	α 0,90	α 0,95	α 0,99	n	r	α 0,90	α 0,95	α 0,99
19	3	0,90	0,95	0,99	21	3	0,90	0,95	0,99	22	19	0,64	0,68	0,75	24	13	0,68	0,73	0,80
	4	0,69	0,78	0,90		4	0,69	0,78	0,90		20	0,61	0,65	0,72		14	0,64	0,68	0,76
	5	0,81	0,86	0,94		5	0,80	0,86	0,94		21	0,64	0,67	0,73		15	0,67	0,71	0,78
	6	0,68	0,75	0,86		6	0,67	0,74	0,85		22	0,61	0,64	0,70		16	0,63	0,67	0,74
	7	0,75	0,81	0,89		7	0,75	0,80	0,89	23	3	0,90	0,95	0,99		17	0,65	0,69	0,76
	8	0,67	0,72	0,82		8	0,66	0,73	0,82		4	0,68	0,77	0,89		18	0,62	0,66	0,73
	9	0,72	0,77	0,85		9	0,72	0,77	0,85		5	0,80	0,86	0,94		19	0,64	0,68	0,75
	10	0,65	0,71	0,80		10	0,65	0,70	0,80		6	0,68	0,76	0,86		20	0,61	0,65	0,72
	11	0,69	0,74	0,82		11	0,69	0,74	0,82		7	0,76	0,82	0,89		21	0,64	0,67	0,73
	12	0,64	0,69	0,77		12	0,64	0,69	0,77		8	0,67	0,73	0,83		22	0,61	0,64	0,71
	13	0,68	0,72	0,80		13	0,68	0,72	0,79		9	0,72	0,78	0,86		23	0,63	0,66	0,72
	14	0,63	0,68	0,76		14	0,63	0,67	0,75		10	0,66	0,71	0,80		24	0,60	0,64	0,69
	15	0,66	0,70	0,78		15	0,66	0,70	0,78		11	0,70	0,75	0,82	25	3	0,90	0,95	0,99
	16	0,62	0,66	0,74		16	0,63	0,67	0,74		12	0,64	0,69	0,78		4	0,69	0,78	0,91
	17	0,65	0,69	0,76		17	0,65	0,69	0,76		13	0,68	0,73	0,80		5	0,81	0,87	0,94
	18	0,61	0,65	0,72		18	0,62	0,66	0,73		14	0,63	0,68	0,76		6	0,68	0,75	0,86
	19	0,64	0,67	0,74		19	0,64	0,68	0,75		15	0,67	0,71	0,78		7	0,75	0,81	0,89
20	3	0,90	0,95	0,99		20	0,61	0,65	0,72		16	0,63	0,67	0,75		8	0,66	0,72	0,82
	4	0,68	0,78	0,90		21	0,63	0,67	0,73		17	0,65	0,69	0,77		9	0,72	0,77	0,85
	5	0,80	0,86	0,91	22	3	0,90	0,95	0,99		18	0,62	0,66	0,73		10	0,65	0,70	0,80
	6	0,67	0,75	0,86		4	0,68	0,77	0,90		19	0,64	0,68	0,75		11	0,70	0,75	0,82
	7	0,75	0,81	0,89		5	0,80	0,86	0,94		20	0,61	0,65	0,72		12	0,64	0,69	0,78
	8	0,66	0,73	0,82		6	0,68	0,75	0,85		21	0,63	0,67	0,73		13	0,68	0,73	0,81
	9	0,72	0,77	0,85		7	0,75	0,81	0,89		22	0,60	0,64	0,70		14	0,63	0,68	0,76
	10	0,65	0,71	0,80		8	0,66	0,72	0,82		23	0,63	0,66	0,72		15	0,66	0,71	0,78
	11	0,69	0,74	0,83		9	0,72	0,77	0,85	24	3	0,90	0,95	0,99		16	0,63	0,67	0,74
	12	0,64	0,69	0,77		10	0,65	0,70	0,80		4	0,69	0,78	0,90		17	0,65	0,69	0,76
	13	0,68	0,72	0,80		11	0,69	0,74	0,82		5	0,81	0,86	0,94		18	0,62	0,66	0,73
	14	0,63	0,68	0,76		12	0,64	0,69	0,78		6	0,68	0,75	0,85		19	0,64	0,68	0,75
	15	0,66	0,71	0,78		13	0,68	0,72	0,80		7	0,75	0,81	0,89		20	0,61	0,65	0,72
	16	0,62	0,67	0,74		14	0,63	0,68	0,75		8	0,67	0,73	0,83		21	0,63	0,67	0,74
	17	0,65	0,69	0,76		15	0,67	0,71	0,78		9	0,72	0,77	0,86		22	0,61	0,64	0,71
	18	0,62	0,66	0,72		16	0,62	0,67	0,74		10	0,66	0,71	0,80		23	0,63	0,66	0,72
	19	0,64	0,68	0,74		17	0,65	0,69	0,76		11	0,70	0,75	0,83		24	0,60	0,63	0,70
	20	0,61	0,65	0,71		18	0,62	0,66	0,73		12	0,64	0,70	0,78		25	0,62	0,65	0,71

$$\sum_{i=1}^{9} \frac{\ln x_{i+1} - \ln x_i}{a_i} = 3{,}929959 + \frac{0{,}693}{1{,}053606} + \ldots + \frac{2{,}398 - 2{,}197}{0{,}28163} = 8{,}901;$$

$$K = \frac{3{,}929959}{8{,}901} = 0{,}441.$$

Из табл. 101 для $n = r = 10$ имеем $K_{0,95}(10{,}10) = 0{,}69$.

Так как $K = 0{,}441 < K_{0,95}(10, 10) = 0{,}69$, нулевая гипотеза не отклоняется. Она включает в себя и гипотезу экспоненциальности распределения, так как экспоненциальное распределение является частным случаем распределения Вейбулла. Рассмотрим случай цензурированных испытаний при $r = 7 < n = 10$. Для этого случая имеем

$$\sum_{i=4}^{6} \frac{\ln x_{i+1} - \ln x_i}{a_i} = 4{,}3402333; \quad \sum_{i=1}^{6} \frac{\ln x_{i+1} - \ln x_i}{a_i} = 6{,}886416;$$

$$K = \frac{4{,}3402333}{6{,}886416} = 0{,}630026.$$

Из табл. 101 для $n = 10$ и $r = 7$ находим $K_{0,95}(7, 10) = 0{,}81$.

Так как $K = 0{,}63 < K_{0,95}(7, 10) = 0{,}81$, нулевая гипотеза не отклоняется и в этой ситуации.

3.3.22. Критерий Дешпанде

В [334] Дешпанде рассмотрел класс критериев экспоненциальности против альтернатив распределений с увеличивающейся интенсивностью отказов, основанных на статистике

$$J = \frac{1}{n(n-1)} \sum_{i \neq j} \varphi(x_i; bx_j), \quad \text{где} \quad \varphi(x_i; bx_j) = \begin{cases} 1, & \text{если } x_i > bx_j; \\ 0, & \text{если } x_i \leqslant bx_j. \end{cases}$$

В условиях, когда нулевая гипотеза справедлива, величина $\sqrt{n\left(J - \dfrac{1}{1+b}\right)}$ распределена асимптотически нормально с нулевым средним и дисперсией [326]

$$\mathbf{D}\left[\sqrt{n}\left(J - \frac{1}{1+b}\right)\right] = \frac{1}{1+2b} + \frac{2b}{(1+b)^2} - \frac{2b}{1+b+b^2} - \frac{2}{(1+b)^2(2+b)}.$$

От выбора постоянной b зависит эффективность критерия против различных альтернатив. Мы воспользуемся рекомендацией авторов работы [326], установивших, что для наиболее распространенной альтернативы вейбулловского распределения следует брать $b = 0{,}44$.

Тогда критерий отклоняет нулевую гипотезу в случае, если

$$\sum_{i \neq j} \varphi(x_i; 0{,}44x_j) \geqslant 0{,}1494\sqrt{n}(n-1)\, u_\alpha + 0{,}694n(n-1).$$

Задача 147. *В условиях задачи* 125 *проверить соответствие эмпирических данных экспоненциальному распределению критерием Дешпанде при* $\alpha = 0{,}95$.

Имеем для различных сочетаний i и j:

$$\varphi(x_1; 0{,}44x_2) = \varphi(1; 0{,}44 \cdot 2) = \varphi(1; 0{,}88) = 1;$$
$$\varphi(x_1; 0{,}44x_3) = \varphi(1; 0{,}44 \cdot 4) = \varphi(1; 1{,}76) = 0;$$
$$\varphi(x_1; 0{,}44x_4) = \ldots = \varphi(x_1; 0{,}44x_{10}) = 0; \quad \varphi(x_2; 0{,}44x_3) = 1;$$
$$\varphi(x_2; 0{,}44x_4) = \ldots = \varphi(x_2; 0{,}44x_{10}) = 0; \quad \varphi(x_3; 0{,}44x_4) = 1;$$

$$\varphi(x_4; 0{,}44x_5) = \varphi(x_4; 0{,}44x_6) = 1; \quad \varphi(x_4; 0{,}44x_7) = \ldots = \varphi(x_4; ,044x_{10}) = 0;$$
$$\varphi(x_5; 0{,}44x_6) = \varphi(x_5; 0{,}44x_7) = 1; \quad \varphi(x_5; 0{,}44x_8) = \ldots = \varphi(x_5; 0{,}44x_{10}) = 0;$$
$$\varphi(x_6; 0{,}44x_7) = \varphi(x_6; 0{,}44x_8) = 1; \quad \varphi(x_6; 0{,}44x_9) = \varphi(x_9; 0{,}44x_{10}) = 0;$$
$$\varphi(x_7; 9{,}44x_8) = \ldots = \varphi(x_7; 0{,}44x_{10}) = 1;$$
$$\varphi(x_8; 0{,}44x_9) = \varphi(x_8; 0{,}44x_{10}) = \varphi(x_9; 0{,}44x_{10}) = 1;$$
$$\sum_{i \neq j} \varphi(x_i; 0{,}44x_j) = 17.$$

В нашем случае $U_{0,95} = 1{,}645$ и $n = 10$. Тогда имеем

$$0{,}1494 \cdot \sqrt{10} \cdot (10 - 1) \cdot 1{,}645 + 0{,}694 \cdot 10 \cdot (10 - 1) = 69{,}454.$$

Так как $\sum_{i \neq j} \varphi(x_i; 0{,}44x_j) = 17 < 69{,}454$, гипотеза экспоненциальности распределения случайных величин не отклоняется.

3.3.23. Критерий Лоулесса

В работах Лоулесса [336], Энгельгардта и Байна [335] рассмотрен критерий проверки гипотезы относительно величины параметра формы гамма-распределения, основанный на статистике

$$W = \frac{\tilde{x}}{\bar{x}},$$

где $\tilde{x} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$; $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ — соответственно геометрическое и арифметическое средние ряда $x_1, x_2, \ldots, x_n$.

Напомним, что при равенстве параметра формы гамма-распределения единице оно переходит в экспоненциальное. Поэтому проверка нулевой гипотезы $H_0\colon k = 1$ против альтернативы $H_1'\colon k > 1$ или $H_1''\colon k < 1$ эквивалентна проверке гипотезы экспоненциальности выборочных данных. Гипотеза экспоненциальности принимается на уровне значимости α, если выборочное значение W находится в интервале $W_1(\alpha) < W < W_2(\alpha)$. Если $W < W_1(\alpha)$, то нулевая гипотеза отклоняется в пользу гипотезы H_1'' (т. е. в пользу гипотезы о распределении с уменьшающейся интенсивностью отказов).

Если $W > W_2(\alpha)$, то нулевая гипотеза отклоняется в пользу альтернативы H_1' (распределение с увеличивающейся интенсивностью отказов).

Статистика W является обобщением ранее рассмотренной нами статистики Бартлетта–Морана (см. раздел 3.3.9). Критические значения $W_1(\alpha)$ и $W_2(\alpha)$ для проверки гипотезы экспоненциальности, заимствованные из [337], приведены в табл. 102.

Задача 148. *В условиях задачи* 125 *проверить соответствие эмпирических данных экспоненциальному распределению критерием Лоулесса при* $\alpha = 0{,}05$.

Имеем

$$\tilde{x} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = (1 \cdot 2 \cdot 4 \cdot 5 \cdot 9 \cdot 11 \cdot 18 \cdot 21 \cdot 29 \cdot 35)^{\frac{1}{10}} = 8{,}28257;$$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{10} x_i = 13{,}5; \quad W = \frac{\tilde{x}}{\bar{x}} = \frac{8{,}28257}{13{,}5} = 0{,}6135.$$

Из табл. 102 для $n = 10$ и $\alpha = 0{,}05$ имеем $W_1 = 0{,}3758$ и $W_2 = 0{,}8210$.

Так как $W_1 = 0{,}3758 < W = 0{,}6135 < W_2 = 0{,}821$, гипотеза экспоненциальности не отклоняется.

Таблица 102

Критические значения $W_1(\alpha)$ и $W_2(\alpha)$ критерия экспоненциальности Лоулесса [337]

n	Уровень значимости α					
	0,10		0,05		0,01	
	W_1	W_2	W_1	W_2	W_1	W_2
2	0,4359	0,9950	0,3122	0,9987	0,1411	0,9999
3	0,3991	0,9574	0,3058	0,9790	0,1670	0,9959
4	0,3966	0,9147	0,3173	0,9477	0,1927	0,9826
5	0,4006	0,8794	0,3299	0,9176	0,2157	0,9646
6	0,4061	0,8512	0,3413	0,8915	0,2344	0,9458
7	0,4115	0,8285	0,3515	0,8695	0,2505	0,9280
8	0,4165	0,8099	0,3601	0,8508	0,2644	0,9115
9	0,4216	0,7944	0,3684	0,8348	0,2763	0,8967
10	0,4261	0,7812	0,3758	0,8210	0,2876	0,8832
11	0,4302	0,7699	0,3818	0,8089	0,2973	0,8710
12	0,4339	0,7601	0,3877	0,7982	0,3059	0,8600
13	0,4373	0,7514	0,3929	0,7887	0,3139	0,8500
14	0,4405	0,7437	0,3977	0,7802	0,3211	0,8408
15	0,4435	0,7368	0,4019	0,7726	0,3275	0,8324
16	0,4463	0,7306	0,4058	0,7656	0,3334	0,8241
17	0,4487	0,7350	0,4094	0,7593	0,3393	0,8175
18	0,4513	0,7199	0,4130	0,7536	0,3446	0,8109
19	0,4535	0,7151	0,4162	0,7482	0,3491	0,8047
20	0,4556	0,7108	0,4194	0,7432	0,3542	0,7989
21	0,4574	0,7070	0,4219	0,7388	0,3580	0,7936
22	0,4595	0,7031	0,4249	0,7343	0,3619	0,7885
23	0,4613	0,6998	0,4273	0,7305	0,3655	0,7838
24	0,4629	0,6964	0,4296	0,7265	0,3690	0,7793
25	0,4644	0,6933	0,4317	0,7230	0,3724	0,7750
26	0,4662	0,6906	0,4341	0,7199	0,3761	0,7710
27	0,4674	0,6878	0,4360	0,7167	0,3786	0,7672
28	0,4688	0,6854	0,4380	0,7137	0,3817	0,7637
29	0,4700	0,6830	0,4398	0,7110	0,3843	0,7602
30	0,4713	0,6808	0,4416	0,7083	0,3868	0,7571

3.4. Критерии согласия для равномерного распределения

Напомним, если $x_1, \ldots, x_n$ — выборка из распределения вероятностей с функцией $F(x)$, то случайная величина $y_i = F(x_i)$ распределена равномерно на интервале $[0,1]$. Поэтому установление равномерности распределения y_i является по существу критерием согласия наблюдаемых данных с теоретическим распределением $F(x)$. Этим объясняется повышенный интерес к поиску простых в вычислительном отношении и эффективных критериев равномерности распределения случайных величин. Можно указать на такие критерии равномерности распределения, как

— критерий Кимбела [338]

$$A = \sum_{i=1}^{n+1} \left(U_i - U_{i-1} - \frac{1}{n+1} \right)^2, \quad U_{n+1} = 1;$$

— критерий Морана [339]

$$B = \sum_{i=1}^{n+1} (U_i - U_{i-1})^2, \quad U_{n+1} = 1;$$

— критерий ω^2 [340]

$$\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(U_i - \frac{2i-1}{n} \right)^2.$$

Здесь U_i — порядковая статистика равномерно распределенной случайной величины. В настоящем разделе мы рассмотрим наиболее эффективные критерии равномерности распределения случайных величин.

3.4.1. Критерий Шермана

Статистика критерия Шермана для проверки равномерности распределения имеет вид [327]

$$\omega_{n=1} = \frac{1}{2} \sum_{i=1}^{n+1} \left| U_i - U_{i-1} - \frac{1}{n+1} \right|, \quad U_0 = 1, \quad U_n = 1.$$

Распределение статистики Шермана рассмотрено в разделе 3.3.15, там же приведены критические значения (табл. 95) и аппроксимации статистики этого критерия.

Задача 149. *Имеется ряд наблюдений над случайной величиной:*

U_i: 0,047; 0,05; 0,15; 0,18; 0,29; 0,48; 0,52; 0,61; 0,72; 0,91.

Необходимо проверить гипотезу равномерности распределения случайной величины U_i *на интервале* $[0,1]$ *критерием Шермана при доверительной вероятности* $\alpha = 0{,}95$.

Находим

$$\omega_n = \frac{1}{2} \cdot \left(\left| 0{,}047 - \frac{1}{11} \right| + \left| 0{,}5 - 0{,}047 - \frac{1}{11} \right| \ldots \left| 0{,}72 - 0{,}61 - \frac{1}{11} \right| + \left| 0{,}91 - 0{,}72 - \frac{1}{11} \right| \right) = \\ = 0{,}245.$$

Из табл. 95 для $n = 10$ имеем $\omega_{10}(0{,}95) = 0{,}473$.

Так как $\omega_n = 0{,}245 < \omega_{10}(0{,}95) = 0{,}473$, гипотеза равномерности не отклоняется.

Рассмотрим нормальную аппроксимацию ω_n-распределения.

Имеем (см. раздел 3.3.15)

$$\mathbf{M}(\omega_n) = \left(\frac{n}{n+1}\right)^{n+1} = \left(\frac{10}{11}\right)^{11} = 0{,}35049;$$

$$\mathbf{D}(\omega_n) = \frac{2 \cdot 10^{12} + 10 \cdot 9^{12}}{12 \cdot 11^{12}} - \left(\frac{10}{11}\right)^{22} = 0{,}005251 \; (\sqrt{\mathbf{D}(\omega_n) = 0{,}07246}).$$

Вычисляем

$$\tilde{\omega}_n = \frac{\omega_n - \mathbf{M}(\omega_n)}{\sqrt{\mathbf{D}(\omega_n)}} = \frac{0{,}245 - 0{,}35049}{0{,}07246} = -1{,}457.$$

Так как $|\tilde{\omega}_n| = 1{,}457 < u_{0,95} = 1{,}645$ (95%-я квантиль стандартного нормального распределения), гипотеза равномерности распределения не отклоняется.

И еще одна аппроксимация:

$$U = \frac{\omega_n - 0{,}3679 \cdot \left(1 - \frac{1}{2n}\right)}{\frac{0{,}243}{\sqrt{n}} \cdot \left(1 - \frac{0{,}605}{n}\right)} = -1{,}4475; \quad U' = -1{,}4475 - \frac{0{,}0995}{\sqrt{10}} \cdot \left(1{,}4475^2 - 1\right) = -1{,}482.$$

Видим, что результаты близки — случайная величина U', имеющая стандартное нормальное распределение, находится в пределах 95%-го интервала, что позволяет принять гипотезу равномерности.

3.4.2. Критерий Морана

Статистика критерия, предложенного Мораном [309], используется в форме

$$M_n = -\sum_{i=1}^{n+1} \ln[(n+1)D_i],$$

где $D_i = U_i - U_{i-1}$, $D_1 = U_1$, $D_{n+1} = 1 - U_n$ $(i = 1, \ldots, n+1)$.

В [231] показано, что случайная величина

$$\tilde{M}_n = \frac{12(n+1)}{7n+8} M_n$$

имеет распределение χ^2 с $f = n$ степенями свободы. Используя аппроксимацию Вилсона–Хилферти (см. раздел 1.1.8), можно получить аппроксимацию

$$z = \frac{\left[\frac{12(n+1)}{7n+8} M_n\right]^{\frac{1}{3}} - \left(1 - \frac{2}{9n}\right)}{\sqrt{\frac{2}{9n}}},$$

где z — случайная величина, имеющая стандартное нормальное распределение.

Гипотеза равномерности принимается, если

$$M_1(\alpha) \leqslant M_n \leqslant M_2(\alpha)\,,$$

где $M_1(\alpha)$ и $M_2(\alpha)$ — критические значения статистики, приведенные в табл. 103.

Для частного случая, когда $U_\imath = F(x_i)$, критерий Морана становится критерием согласия эмпирического распределения с теоретическим $F(x)$. В случае их совпадения случайная величина U_i должна быть распределена равномерно на интервале $[0, 1]$.

Таблица 103

Критические значения $M_1(\alpha)$ и $M_2(\alpha)$ статистики Морана [341]

n	Доверительная вероятность α					
	0,90		0,95		0,99	
	M_1	M_2	M_1	M_2	M_1	M_2
2	0,130	2,756	0,064	3,554	0,012	5,370
3	0,356	3,699	0,215	4,592	0,070	6,579
4	0,643	4,573	0,430	5,546	0,180	7,675
5	0,967	5,408	0,689	6,449	0,334	8,705
6	1,317	6,214	0,979	7,319	0,523	9,688
7	1,687	7,001	1,292	8,162	0,741	10,637
8	2,072	7,772	1,625	8,987	0,982	11,559
9	2,469	8,531	1,973	9,795	1,242	12,460
10	2,876	9,279	2,333	10,591	1,519	13,343
11	3,292	10,018	2,704	11,375	1,810	14,210
12	3,715	10,750	3,085	12,150	2,113	15,064
13	4,145	11,475	3,479	12,916	2,428	15,907
14	4,581	12,194	3,870	13,675	2,752	16,740
15	5,021	12,909	4,272	14,428	3,085	17,563
16	5,466	13,618	4,681	15,174	3,425	18,378
17	5,916	14,324	5,094	15,916	3,774	19,186
18	6,369	15,025	5,513	16,652	4,128	19,987
19	6,825	15,721	5,936	17,384	4,489	20,782
20	7,285	16,418	6,362	18,112	4,855	21,571

Если параметры теоретического распределения известны заранее, то критерий Морана используется в форме [342]

$$M_n = -\sum_{i=1}^{n+1} \ln D_i,$$

где $D_i = U_i - U_{i-1}$, $i = 1, \ldots, n+1$, $U_0 = 0$, $U_{n+1} = 1$.

Равномерность распределения U_i на интервале $[0,1]$, а следовательно, и допустимость использования в качестве вероятностной модели распределения функции $F(x)$, отклоняется, если [342]

$$\frac{M_n - c_1}{c_2} > \chi^2_\alpha(n)\,,$$

где $c_1 = (n+1)[\ln(n+1) + 0{,}5722] - \frac{1}{2} - \frac{1}{12(n+1)} - \sqrt{\frac{n}{2}}(n+1)\left(\frac{\pi^2}{6} - 1\right) - \frac{1}{2} - \frac{1}{6(n+1)}$;

$$c_2 = \frac{1}{\sqrt{2n}}(n+1)\left(\frac{\pi^2}{6} - 1\right) - \frac{1}{2} - \frac{1}{6(n+1)};$$

$\chi^2_\alpha(n)$ — α-квантиль стандартного нормального распределения с $f = n$ степенями свободы.

Если параметры распределения $F(x)$ определяются по выборке, то критерий имеет вид

$$\frac{M_n + \frac{k}{2} - c_1}{c_2} > \chi^2_\alpha(n)\,,$$

где k — число определяемых по выборке параметров.

Эти аппроксимации, предложенные Ченгом и Стефенсом [342], дают весьма удовлетворительные результаты уже при $n > 5$.

Задача 150. *Для данных задачи 149 проверить гипотезу равномерности распределения случайных величин критерием Морана при доверительной вероятности* $\alpha = 0{,}95$.

Имеем

$$D_1 = U_1 = 0{,}047;\quad D_2 = 0{,}05 - 0{,}047 = 0{,}003;\quad D_3 = 0{,}15 - 0{,}05 = 0{,}10;$$
$$D_4 = 0{,}18 - 0{,}15 = 0{,}03;\quad D_5 = 0{,}11;\quad D_6 = 0{,}19;\quad D_7 = 0{,}04;\quad D_8 = 0{,}04;$$
$$D_9 = 0{,}09;\quad D_{10} = 11;\quad D_{11} = 0{,}09.$$

Тогда

$$M_n = \sum_{i=1}^{11} \ln\left[(n+1)\cdot D_i\right] = -11\ln 11 - \sum_{i=1}^{11} \ln D_i =$$
$$= 26{,}3768 - (\ln 0{,}047 + \ln 0{,}003 + \ldots + \ln 0{,}19 + \ln 0{,}09) = 4{,}0698.$$

Из табл. 103 находим $M_1 = 2{,}333$ и $M_2 = 10{,}591$.

Так как $M_1 = 2{,}333 < M_n = 4{,}0698 < M_2 = 10{,}591$, гипотеза равномерности не отклоняется.

Воспользуемся теперь *нормальной аппроксимацией* и вычислим

$$z = \frac{\left(\dfrac{12\cdot 11}{780}\cdot 4{,}0698\right)^{\frac{1}{3}} - \left(1 - \dfrac{2}{90}\right)}{\sqrt{\dfrac{2}{90}}} = -0{,}635,$$

что находится в диапазоне $z_{0,05} \div z_{0,95} = -1{,}645 \div +1{,}645$.

Случайная величина $\tilde{M}_n = \dfrac{12\cdot(n+1)}{7n+8}\cdot M_n = \dfrac{12\cdot 11}{78}\cdot 4{,}0698 = 6{,}887$ имеет χ^2-распределение с $f = 10$ степенями свободы. При $\alpha = 0{,}95$ значение $\chi^2_{0,95}(10) = 18{,}307$ (см. табл. 55), что больше $\tilde{M}_n$ и, следовательно, не отклоняет гипотезу равномерности распределения.

Применим теперь *аппроксимацию Ченга–Стефенса.*

Имеем $M_n = \sum\limits_{i=1}^{11} \ln D_i = 30{,}4467$. Далее вычисляем

$$c_1 = (10+1)\cdot(\ln 11 + 0{,}57722) - \frac{1}{2} - \frac{1}{12\cdot 11} - \sqrt{5}\cdot 11\cdot\left(\frac{\pi^2}{6} - 1\right) - \frac{1}{2} - \frac{1}{6\cdot 11} = 15{,}84026;$$

$$c_2 = \frac{1}{\sqrt{2\cdot 10}}\cdot 11\cdot\left(\frac{\pi^2}{6} - 1\right) - \frac{1}{2} - \frac{1}{6\cdot 11} = 1{,}071176.$$

Тогда $\dfrac{M_n - c_1}{c_2} = \dfrac{30{,}4467 - 15{,}84026}{1{,}071176} = 13{,}636 < \chi^2_{0,95}(10) = 18{,}307$, и этот критерий не отклоняет нулевую гипотезу.

3.4.3. Критерий Ченга–Спиринга

Критерий равномерности распределения, аналогичный критерию Шапиро–Уилки (см. разделы 3.2.2.1 и 3.3.1), предложен в [343]. Его статистика имеет вид

$$W_p = \frac{\left[(x_n - x_1)\,\dfrac{n+1}{n-1}\right]^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2},$$

где $x_n - x_1$ — выборочный размах.

Всегда

$$\frac{2(n+1)^2}{(n-1)^2} \geqslant W_p \geqslant \frac{4(n+1)^2}{n(n-1)^2} \quad (n = 2, 4, \ldots, 2k);$$

$$\frac{2(n+1)^2}{(n-1)^2} \geqslant W_p \geqslant \frac{4n(n+1)}{(n-1)^3} \quad (n = 1, 3, \ldots, 2k-1)$$

при справедливости нулевой гипотезы (подробнее об этом см. в [211]).

Гипотеза равномерности отклоняется, если $W_1(\alpha) \leqslant W_p \leqslant W_2(\alpha)$, где $W_1(\alpha)$ и $W_2(\alpha)$ — критические значения при доверительной вероятности α, приведенные в табл. 104.

Таблица 104

Критические значения $W_1(\alpha)$ и $W_2(\alpha)$ критерия равномерности Ченга–Спиринга [143]

n	Доверительная вероятность α						n	Доверительная вероятность α					
	0,90		0,95		0,99			0,90		0,95		0,99	
	W_1	W_2	W_1	W_2	W_1	W_2		W_1	W_2	W_1	W_2	W_1	W_2
3	6,30	7,97	6,15	7,99	6,03	8,00	16	0,65	1,03	0,62	1,11	0,56	1,31
4	3,74	5,31	3,44	5,43	3,08	5,53	17	0,61	0,97	0,58	1,04	0,53	1,20
5	2,58	4,02	2,42	4,18	2,40	4,39	18	0,58	0,90	0,55	0,98	0,50	1,12
6	2,00	3,23	1,88	3,41	1,71	3,67	19	0,55	0,85	0,52	0,92	0,47	1,05
7	1,64	2,68	1,54	2,85	1,39	3,13	20	0,52	0,80	0,50	0,86	0,45	0,98
8	1,40	2,32	1,32	2,47	1,18	2,77	21	0,50	0,76	0,47	0,81	0,43	0,92
9	1,22	2,04	1,15	2,19	1,04	2,46	22	0,47	0,72	0,45	0,78	0,41	0,89
10	1,08	1,79	1,02	1,92	0,91	2,18	23	0,45	0,68	0,43	0,73	0,40	0,84
11	0,97	1,60	0,92	1,73	0,83	1,99	24	0,44	0,65	0,42	0,70	0,38	0,80
12	0,89	1,45	0,84	1,57	0,76	1,79	25	0,42	0,62	0,40	0,67	0,36	0,76
13	0,81	1,31	0,77	1,42	0,69	1,64	30	0,35	0,51	0,33	0,54	0,31	0,61
14	0,75	1,22	0,71	1,31	0,64	1,52	40	0,26	0,37	0,25	0,39	0,23	0,44
15	0,70	1,12	0,66	1,21	0,60	1,39	50	0,21	0,29	0,20	0,30	0,19	0,33

Задача 151. *Для данных задачи* 149 *проверить гипотезу равномерности распределения случайной величины критерием Ченга–Спиринга при* $\alpha = 0{,}95$.

Имеем $x_n = 0{,}91$, $x_1 = 0{,}047$ и $\sum\limits_{i=1}^{10}(x_i - \bar{x})^2 = 0{,}7973241$.

Тогда $W_p = \dfrac{\left[(0{,}91 - 0{,}047) \cdot \dfrac{11}{9}\right]^2}{0{,}7973241} = 1{,}395$. Из табл. 104 для $\alpha = 0{,}95$ и $n = 10$ находим $W_1(0{,}95) = 1{,}02$ и $W_2(0{,}95) = 1{,}92$.

Так как $W_1(\alpha) = 1{,}102 < W_p = 1{,}395 < W_2(\alpha) = 1{,}92$, гипотеза равномерности распределения совокупности случайных величин не отклоняется.

3.4.4. Критерий Саркади–Косика

В [344] рассмотрен новый, весьма эффективный критерий равномерности на $[0, 1]$, основанный на модификации критерия Ватсона (см. раздел 1.2.5), со статистикой

$$J = n^2 \sum_{i=1}^{n} d_i^2 - n\left(\sum_{i=1}^{n} d_i\right)^2,$$

где $d_i = \dfrac{x_i - \dfrac{i}{n+1}}{i\,(n-i+1)}$ и $x_i \leqslant x_2 \leqslant \ldots \leqslant x_n$.

Распределение признается не противоречащим равномерному на уровне значимости α, если $J < J(\alpha)$, где $J(\alpha)$ — критическое значение, приведенное в табл. 105.

Таблица 105

Критические значения $J(\alpha)$ критерия Саркади–Косика [344]

n	Доверительная вероятность α		n	Доверительная вероятность α	
	0,05	0,10		0,05	0,10
5	0,499	0,408	40	1,040	0,836
10	0,741	0,599	50	1,020	0,823
15	0,881	0,695	60	1,060	0,862
20	0,931	0,748	70	1,070	0,870
25	0,973	0,782	80	1,070	0,876
30	1,000	0,806	90	1,080	0,880
35	1,020	0,823	100	1,080	0,883

Критерий обладает высокой мощностью против большого спектра альтернатив.

Задача 152. *Для данных задачи* 149 *проверить гипотезу равномерности распределения случайной величины критерием Саркади–Косика на уровне значимости* $\alpha = 0{,}05$.

Вычисления статистики критерия приведены в таблице:

i	x_i	$\dfrac{i}{n+1}$	$x_i - \dfrac{i}{n+1}$	$i(n-i+1)$	d_i	d_i^2
1	0,047	0,0909	−0,0439	10	−0,004390	$1{,}927 \cdot 10^{-5}$
2	0,050	0,1818	−0,1318	18	−0,007322	$5{,}631 \cdot 10^{-5}$
3	0,150	0,2727	−0,1227	24	−0,005110	$2{,}614 \cdot 10^{-5}$
4	0,180	0,3636	−0,1836	28	−0,006557	$4{,}300 \cdot 10^{-5}$
5	0,290	0,4545	−0,1645	30	−0,005480	$3{,}010 \cdot 10^{-5}$
6	0,480	0,5454	−0,0654	30	−0,002180	$4{,}750 \cdot 10^{-5}$
7	0,520	0,6363	−0,1163	28	−0,004150	$1{,}720 \cdot 10^{-5}$
8	0,610	0,7272	−0,1172	24	−0,004880	$2{,}380 \cdot 10^{-5}$
9	0,720	0,8181	−0,0981	18	−0,005450	$2{,}970 \cdot 10^{-5}$
10	0,910	0,9090	0,0010	10	0,000100	10^{-8}

Вычисляем статистику

$$J = 100 \cdot 2{,}5028 \cdot 10^{-4} - 10 \cdot (0{,}045419)^2 = 4{,}399 \cdot 10^{-3}.$$

Легко видеть, что $J = 4{,}399 \cdot 10^{-3} < J(\alpha) = 0{,}741$, что не отклоняет гипотезу равномерности.

3.4.5. Энтропийный критерий Дудевича–ван дер Мюлена

По аналогии с критерием нормальности Васичека [247] (см. раздел 3.2.2.2) в [345] рассмотрен критерий для оценки равномерности распределения.

Пусть $x_1 \leqslant \ldots \leqslant x_n$ — выборка случайных величин. Ее выборочная энтропия равна [247]

$$H(m,n) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\log_2 \frac{n}{2m}(x_{i+m} - x_{i-m})\right\},$$

m — целое число $\leqslant \dfrac{n}{2}$ (используем $x_i = x_1$ при $i < 1$ и $x_i = x_n$ при $i > n$).

Распределение случайных величин признается равномерным, если $H(m,n) \leqslant H_\alpha(m,n)$, где $H_\alpha(m,n)$ — критическое значение, приведенное в табл. 106.

Таблица 106

Критические значения $H_\alpha(m,n)$ статистики Дудевича–ван дер Мюлена [346]

n	m	Уровень значимости α			n	m	Уровень значимости α		
		0,01	0,05	0,10			0,01	0,05	0,10
10	1	1,228	0,974	0,845	40	8	0,327	0,274	0,249
	2	0,979	0,766	0,659		9	0,339	0,285	0,260
	3	0,951	0,741	0,645		10	0,350	0,297	0,272
	4	0,977	0,774	0,681		15	0,415	0,367	0,316
20	1	0,787	0,658	0,585	50	1	0,532	0,458	0,422
	2	0,588	0,466	0,412		2	0,342	0,282	0,256
	3	0,536	0,429	0,378		3	0,284	0,236	0,213
	4	0,536	0,428	0,378		4	0,264	0,219	0,198
	5	0,543	0,438	0,391		5	0,257	0,214	0,195
	6	0,557	0,458	0,410		6	0,257	0,215	0,196
	7	0,581	0,481	0,433		7	0,260	0,219	0,201
	8	0,605	0,509	0,460		8	0,266	0,226	0,207
	9	0,632	0,536	0,489		9	0,274	0,234	0,216
30	1	0,665	0,544	0,496		10	0,283	0,243	0,224
	2	0,540	0,367	0,327		15	0,337	0,295	0,276
	3	0,402	0,324	0,287		20	0,397	0,355	0,336
	4	0,384	0,312	0,279	100	1	0,430	0,381	0,360
	5	0,384	0,314	0,282		2	0,353	0,219	0,204
	6	0,394	0,322	0,291		3	0,199	0,170	0,158
	7	0,407	0,335	0,304		4	0,175	0,150	0,138
	8	0,420	0,350	0,319		5	0,165	0,140	0,128
	9	0,440	0,366	0,336		6	0,160	0,135	0,123
	10	0,459	0,383	0,354		7	0,157	0,134	0,122
40	1	0,569	0,491	0,450		8	0,157	0,134	0,123
	2	0,379	0,315	0,283		9	0,159	0,136	0,124
	3	0,330	0,270	0,242		10	0,160	0,138	0,128
	4	0,316	0,254	0,229		15	0,180	0,159	0,149
	5	0,311	0,253	0,228		20	0,205	0,185	0,175
	6	0,313	0,257	0,231		30	0,262	0,243	0,233
	7	0,318	0,265	0,239		40	0,325	0,314	0,306

Для $n > 20$ полезна аппроксимация

$$\tilde{H}(m,n) = \sqrt{\frac{n}{3m}} \times \left\{ \sum_{i=0}^{n-m+1} \log_2(x_{i+m} - x_i) + (n+2-m)[\log_2(n+1) + 0{,}5772 - R(1,m-1)] \right\},$$

где

$$R(1,m-1) = \begin{cases} 0 & \text{при } m = 1 \ (x_0 = 0; x_{m+1} = 1); \\ \sum\limits_{j=1}^{m-1} \dfrac{1}{m-j} & \text{при } m \geqslant 2. \end{cases}$$

Статистика $\tilde{H}(m,n)$ имеет стандартное нормальное распределение. Следует помнить, что $H(m,n)$ всегда отрицательно и $\log_2 a = 3{,}3219 \lg a$.

Задача 153. *Для данных задачи 149 проверить гипотезу равномерности распределения случайной величины критерием Дудевича–ван дер Мюлена на уровне значимости* $\alpha = 0{,}05$.

Выбираем $m = 2$ и вычисляем статистику критерия

$$H(2,10) = \frac{1}{10} \cdot \left\{ \log_2 \frac{10}{4} \cdot (x_3 - x_1) + \ldots + \log_2 \frac{10}{4} \cdot (x_{10} - x_8) \right\} =$$
$$= 0{,}1 \cdot (\log_2 0{,}82 + \log_2 0{,}3325 + \ldots + \log_2 0{,}75) = 0{,}20914.$$

Из табл. 106 для $n = 10$, $m = 2$ и $\alpha = 0{,}05$ имеем $H_{0,05}(2,10) = 0{,}766$.

Так как $H(2,10) = 0{,}209 < H_{0,05}(2,10) = 0{,}766$, гипотеза равномерности распределения вероятностей случайной величины не отклоняется.

3.4.6. Критерий Хегази–Грина

По аналогии с критерием нормальности Хегази–Грина (см. раздел 3.2.2.3) авторы работы [249] рассмотрели и критерий равномерности на основе статистик

$$T = \frac{1}{n} \sum_{i=1}^{n} |x_i - \eta_i| \quad \text{и} \quad T_2 = \frac{1}{n} \sum_{i=1}^{n} X_i (x_i - \eta_i)^2,$$

где x_i — порядковые статистики выборки; $\eta_i = \mathbf{M}(x_i)$ — математическое ожидание i-й порядковой статистики.

Критерий применим и для случайных величин, распределенных на любом интервале $[a, b]$. В этом случае следует перейти к переменной $y_i = \dfrac{x_i - x_1}{x_n - x_1}$ $(i = 1, \ldots, n)$ и объем выборки n заменить на $(n - 2)$.

В конечном итоге предлагаются статистики

$$T_1 = \frac{1}{m} \sum_{i=1}^{m} \left| x_i - \frac{i}{m+1} \right| \quad \text{и} \quad T_2 = \frac{1}{m} \sum_{i=1}^{m} \left(x_i - \frac{i}{m+1} \right)^2,$$

$$\text{где } m = \begin{cases} n & \text{для переменной } x_i \in [0{,}1]\,; \\ n-2 & \text{для переменной } y_i = \dfrac{x_i - x_1}{x_n - x_1}. \end{cases}$$

В [249] рассмотрены также модифицированные статистики, которые вместо математического ожидания i-й порядковой статистики $\eta_i = \dfrac{i}{m+1}$ используют его модальное значение $\xi_i = \dfrac{i-1}{m-1}$ (в силу несимметричности распределения порядковых статистик).

Критерии в этом случае вычисляются по формулам

$$T_1^* = \frac{1}{m} \sum_{i=1}^{m} \left| x_i - \frac{i-1}{m-1} \right|; \quad T_2^* = \frac{1}{m} \sum_{i=1}^{m} \left(x_i - \frac{i-1}{m-1} \right)^2.$$

Гипотеза равномерности распределения случайных величин x_i принимается на уровне значимости α, если

$$T_1(T_1^*) < T_1(\alpha)(T_1^*(\alpha)) \quad \text{или} \quad T_2(T_2^*) < T_2(\alpha)(T_2^*(\alpha)),$$

где $T_1(\alpha)$, $T_1^*(\alpha)$, $T_2(\alpha)$, $T_2^*(\alpha)$ — критические значения соответствующих статистик, приведенные в табл. 107.

Таблица 107

Критические значения $T_1(\alpha)$, $T_1^*(\alpha)$, $T_2(\alpha)$ и $T_2^*(\alpha)$ критерия равномерности Хегази–Грина [249]

n	Уровень значимости α							
	0,01				0,05			
	T_1	T_1^*	T_2	T_2^*	T_1	T_1^*	T_2	T_2^*
3	0,3893	0,4277	0,1744	0,2807	0,3254	0,3778	0,1189	0,2059
8	0,2590	0,2674	0,0822	0,0998	0,2030	0,2142	0,0522	0,0641
18	0,1724	0,1764	0,0376	0,0421	0,1360	0,1399	0,0240	0,0268
38	0,1225	0,1224	0,0191	0,0188	0,0950	0,0949	0,0119	0,0123
78	0,0853	0,0843	0,0096	0,0095	0,0664	0,0663	0,0059	0,0059

Для промежуточных значений n применима аппроксимация

$$T(\alpha) = a + \frac{b}{\sqrt{n}} + \frac{c}{n},$$

коэффициенты a, b и c которой для различных статистик и уровней значимости приведены в таблице:

Статистика	Уровень значимости α					
	0,01			0,05		
	a	b	c	a	b	c
$T_1(\alpha)$	−0,0070	0,8373	−0,2500	0,0003	0,5876	−0,0425
$T_1^*(\alpha)$	−0,0090	0,7949	−0,0782	0,0064	0,5066	0,2364
$T_2(\alpha)$	−0,0148	0,1701	0,2745	−0,0068	0,0783	0,2419
$T_2^*(\alpha)$	0,0047	−0,0607	0,9330	0,0214	−0,1395	0,8212

Мощность критерия достаточно велика по сравнению с другими критериями.

Задача 154. *Проверить гипотезу равномерности распределения случайной величины x_i на интервале $[0,1]$ для данных задачи 149 критериями Хегази–Грина на уровне значимости $\alpha = 0{,}05$.*

Имеем

$$T_1 = \frac{1}{n} \cdot \sum_{i=1}^{10} \left| x_i - \frac{i}{n+1} \right| = 0{,}1 \cdot (0{,}0439 + 0{,}1318 + \ldots + 0{,}0982 + 0{,}0083) = 0{,}1051;$$

$$T_2 = \frac{1}{10} \cdot \sum_{i=1}^{10} \left(x_i - \frac{i}{n+1} \right)^2 = 0{,}1 \cdot \left(1{,}9272 \cdot 10\sqrt{-3} + 0{,}01737 + \ldots + 6{,}889 \cdot 10^{-5}\right) = 0{,}0136;$$

$$T_1^* = \frac{1}{10} \cdot \sum_{i=1}^{10} \left| x_i - \frac{i-1}{n-1} \right| = 0{,}1 \cdot (0{,}047 + 0{,}1722 + \ldots + 0{,}16888 + 0{,}09) = 0{,}1248;$$

$$T_2^* = \frac{1}{10} \cdot \sum_{i=1}^{10} \left(x_i - \frac{i-1}{n-1} \right)^2 = 0{,}01764.$$

Из табл. 107 для $\alpha = 0{,}05$ и $n = 10$ находим (аппроксимацией)

$$T_1(\alpha) = 0{,}20; \quad T_1^*(\alpha) = 0{,}21; \quad T_2(\alpha) = 0{,}05; \quad T_2^*(\alpha) = 0{,}06.$$

Более точные критические значения получим, воспользовавшись аппроксимациями:

$$T_1(0{,}05) = 0{,}0003 + \frac{0{,}5876}{\sqrt{10}} - \frac{0{,}0425}{10} = 0{,}182;$$

$$T_1^*(0{,}05) = 0{,}0064 + \frac{0{,}5066}{\sqrt{10}} + \frac{0{,}2364}{10} = 0{,}190;$$

$$T_2(0{,}05) = -0{,}0068 + \frac{0{,}0783}{\sqrt{10}} + \frac{0{,}2419}{10} = 0{,}042;$$

$$T_2^*(0{,}05) = 0{,}0124 - \frac{0{,}1395}{\sqrt{10}} + \frac{0{,}8212}{10} = 0{,}050.$$

Так как

$$T_1 = 0{,}1052 < T_1(0{,}05) = 0{,}182; \quad T_1^* = 0{,}1248 < T_1^*(0{,}05) = 0{,}190;$$
$$T_2 = 0{,}0316 < T(0{,}05) = 0{,}042; \quad T_2^* = 0{,}017 < T_2^*(0{,}05) = 0{,}050,$$

то всеми критериями гипотеза равномерности не отклоняется.

3.4.7. Критерий Янга

В [347] изучен критерий для проверки распределения точек на отрезке. Пусть на отрезке длиной l с началом в нуле расположено n точек с координатами $x_1 \leqslant \ldots \leqslant x_n$. Определим переменные

$$D_1 = x_1, \quad D_i = x_i - x_{i-1}, \quad D_{n+1} = 1 - x_n.$$

Впервые критерии для проверки равномерности, основанные на D-статистиках, рассмотрены в [348, 349], однако пригодные аппроксимации получены в [347].

Статистика критерия Янга имеет вид

$$M = \frac{1}{l}\sum_{i=1}^{n}\min(D_i, D_{i+1}),$$

Очевидно, что $0 \leqslant M \leqslant \dfrac{n}{n+1}$. При $l = 1$ имеет место $M = \sum\limits_{i=1}^{n}\min(D_i, D_{i+1})$.

Если l неизвестно, то используется статистика

$$M^* = \frac{1}{x_n - x_1}\sum_{i=2}^{n-1}\min(D_i, D_{i+1}),$$

распределение которой совпадает с распределением M-статистики при замене n на $(n-2)$.

Гипотеза равномерности не отклоняется, если

$$M_1(\alpha) \leqslant M \leqslant M_2(\alpha)\,,$$

где $M_1(\alpha)$ и $M_2(\alpha)$ — критические значения, приведенные в табл. 108.

При $n \geqslant 15$ можно использовать тот факт, что статистика

$$\tilde{M} = 2(n+1)M\sqrt{\frac{3}{2n-1}} - n\sqrt{\frac{3}{2n-1}}$$

имеет стандартное нормальное распределение. Если u_α — α-квантиль стандартного нормального распределения, то

$$M_{1(2)}(\alpha) = \frac{n}{2(n+1)} \pm u_{\frac{1-\alpha}{2}}\frac{1}{2(n+1)}\sqrt{\frac{2n-1}{3}}.$$

Напомним, что

$$u_{\frac{1-\alpha}{2}} = -1{,}645 \text{ при } \alpha = 0{,}90; \quad u_{\frac{1-\alpha}{2}} = -1{,}96 \text{ при } \alpha = 0{,}95$$
$$\text{и} \quad u_{\frac{1-\alpha}{2}} = -2{,}58 \quad \text{при } \alpha = 0{,}99.$$

Таблица 108

Критические значения $M_1(\alpha)$, $M_2(\alpha)$ статистики Янга [347]

n	Доверительная вероятность α					
	0,90		0,95		0,99	
	M_1	M_2	M_1	M_2	M_1	M_2
2	0,090	0,545	0,043	0,581	0,010	0,545
3	0,151	0,589	0,104	0,627	0,046	0,589
4	0,192	0,602	0,142	0,646	0,073	0,602
5	0,220	0,610	0,170	0,650	0,100	0,610
10	0,300	0,600	0,260	0,640	0,190	0,600
15	0,340	0,590	0,310	0,620	0,240	0,590
20	0,370	0,590	0,330	0,610	0,270	0,590
30	0,390	0,570	0,360	0,600	0,310	0,570
40	0,400	0,570	0,380	0,590	0,340	0,570

Более точно распределение M-статистики аппроксимируется бета-распределением (см. раздел 1.1.7) с параметрами

$$a = b = \frac{3n^2}{2(2n-1)} - \frac{1}{2}.$$

Задача 155. *Для данных задачи* 149 *проверить гипотезу равномерности распределения случайной величины критерием Янга при уровне достоверности* $\alpha = 0{,}95$.

В нашем случае $l = 1$. Имеем ряд

$$D_1 = 0{,}047; \quad D_2 = x_2 - x_1 = 0{,}05 - 0{,}047 = 0{,}003;$$

$$D_3 = 0{,}15 - 0{,}05 = 0{,}10; \quad D_4 = 0{,}18 - 0{,}15 = 0{,}03; \quad D_5 = 0{,}29 - 0{,}18 = 0{,}11;$$

$$D_6 = 0{,}48 - 0{,}29 = 0{,}19; \quad D_7 = 0{,}52 - 0{,}48 = 0{,}04; \quad D_8 = 0{,}61 - 0{,}52 = 0{,}09;$$

$$D_9 = 0{,}72 - 0{,}61 = 0{,}11; \quad D_{10} = 0{,}91 - 0{,}72 = 0{,}19; \quad D_{11} = 1 - 0{,}91 = 0{,}09,$$

из которого следует ряд значений:

$$M_1 = \min(D_1, D_2) = 0{,}003; \quad M_2 = \min(D_2, D_3) = 0{,}003; \quad M_3 = \min(D_3, D_4) = 0{,}03;$$

$$M_4 = \min(D_4, D_5) = 0{,}03; \quad M_5 = \min(D_5, D_6) = 0{,}11; \quad M_6 = \min(D_6, D_7) = 0{,}04;$$

$$M_7 = \min(D_7, D_8) = 0{,}04; \quad M_8 = \min(D_8, D_9) = 0{,}09; \quad M_9 = \min(D_9, D_{10}) = 0{,}11;$$

$$M_{10} = \min(D_{10}, D_{11}) = 0{,}09.$$

Тогда $M = \sum_{i=1}^{10} M_i = 0{,}546$. Из табл. 108 для $n = 10$ и $\alpha = 0{,}95$ имеем $M_1 = 0{,}26$ и $M_2 = 0{,}64$. Так как $M_1 = 0{,}26 < M = 0{,}546 < M_2 = 0{,}64$, гипотеза равномерности не отклоняется.

Если воспользоваться нормальной аппроксимацией, то

$$M_1(0.95) = \frac{10}{2 \cdot 11} - 1{,}645 \cdot \left(\frac{19}{3}\right)^{\frac{1}{2}} \cdot \frac{1}{2 \cdot 11} = 0{,}266;$$

$$M_2(0{,}95) = \frac{10}{2 \cdot 11} + 1{,}645\left(\frac{19}{3}\right)^{\frac{1}{2}} \cdot \frac{1}{2 \cdot 11} = 0{,}642.$$

Видно, что полученные значения близки к табличным.

3.4.8. Критерии типа Колмогорова–Смирнова

Ранее были рассмотрены критерии типа Колмогорова–Смирнова, основанные на сравнении теоретической и эмпирической функций распределения (см. разделы 3.1.2.1 — критерий Колмогорова–Смирнова; 3.1.2.6 — критерий Купера, как общие критерии согласия; 3.2.1.2 — в задачах проверки нормальности распределения; 3.3.2 — в задачах проверки экспоненциальности распределения).

Приведем модифицированные формы критериев Колмогорова–Смирнова (D^+, D^-, D) и Купера (V) для задачи проверки равномерности распределения ($U_1 < U_2 < \ldots < U_n$):

$$D^+ = \max_i\left(U_i - \frac{i}{n+1}\right); \quad D^- = \max_i\left(\frac{i}{n+1} - U_i\right); \quad D = \max(D^+, D^-);$$
$$V = D^+ + D^-.$$

Распределения указанных статистик быстро сходится к предельному, если использовать их модификации:

$$\tilde{D}^+ = \left(D^+ + \frac{0{,}4}{n}\right)\left(\sqrt{n} + 0{,}2 + \frac{0{,}68}{\sqrt{n}}\right); \quad \tilde{D}^- = \left(D^- + \frac{0{,}4}{n}\right)\left(\sqrt{n} + 0{,}2 + \frac{0{,}68}{\sqrt{n}}\right);$$
$$\tilde{D} = \left(D + \frac{0{,}4}{n}\right)\left(\sqrt{n} + 0{,}2 + \frac{0{,}68}{\sqrt{n}}\right); \quad \tilde{V} = \left(V - \frac{1}{n-1}\right)\left(\sqrt{n+1} + 0{,}1555 + \frac{0{,}24}{\sqrt{n+1}}\right).$$

Критические значения для модифицированных статистик $\tilde{D}^+$, $\tilde{D}^-$, $\tilde{D}$, $\tilde{V}$ приведены в табл. 109.

Таблица 109

Критические значения $\tilde{D}^+$, $\tilde{D}^-$, $\tilde{D}$, $\tilde{V}$ критериев равномерности [231]

n	Уровень значимости α				
	0,01	0,025	0,05	0,10	0,15
$\tilde{D}^+$	1,518	1,358	1,224	1,073	0,973
$\tilde{D}^-$	1,518	1,358	1,224	1,073	0,973
$\tilde{D}$	1,628	1,480	1,358	1,224	1,138
$\tilde{V}$	2,001	1,852	1,747	1,620	1,537

Задача 156. *Для данных задачи 149 проверить гипотезу равномерности распределения случайной величины критериями типа Колмогорова–Смирнова на уровне значимости $\alpha = 0{,}05$.*

Результаты вычислений представлены в таблице:

i	U_i	$\frac{i}{n+1}$	$U_i - \frac{i}{n+1}$	$\frac{i}{n+1} - U_i$
1	0,047	0,1111	−0,0641	0,0641
2	0,050	0,1818	−0,0404	0,1318
3	0,150	0,2727	−0,1227	0,1227
4	0,180	0,3636	−0,1836	0,1836
5	0,290	0,4545	−0,1645	0,1645
6	0,480	0,5454	−0,0654	0,0654
7	0,520	0,6363	−0,1163	0,1163
8	0,610	0,7272	−0,1172	0,1172
9	0,720	0,8181	−0,0981	0,0981
10	0,910	0,9090	0,0010	−0,0010

Вычисляем

$$D^+ = 0{,}001; \quad D^- = 0{,}1836; \quad D = 0{,}1836; \quad V = D^+ + D^- = 0{,}1846;$$

$$\tilde{D}^+ = \left(0{,}001 + \frac{0{,}4}{10}\right) \cdot \left(\sqrt{10} + 0{,}2 + \frac{0{,}68}{\sqrt{10}}\right) = 0{,}1466;$$

$$\tilde{D}^- = \left(0{,}1836 + \frac{0{,}4}{10}\right) \cdot \left(\sqrt{10} + 0{,}2 + \frac{0{,}68}{\sqrt{10}}\right) = 0{,}799;$$

$$\tilde{D} = \left(0{,}1836 + \frac{0{,}4}{10}\right) \cdot \left(\sqrt{10} + 0{,}2 + \frac{0{,}68}{\sqrt{10}}\right) = 0{,}799;$$

$$\tilde{V} = \left(0{,}1846 - \frac{1}{9}\right) \cdot \left(\sqrt{11} + 0{,}1555 + \frac{0{,}24}{\sqrt{11}}\right) = 0{,}264.$$

Из табл. 109 для $\alpha = 0{,}05$ находим $\tilde{D} = 1{,}358$ и $\tilde{V} = 1{,}747$.

Так как $\tilde{D} = 0{,}799 < \tilde{D}(0{,}05) = 1{,}358$ и $\tilde{V} = 0{,}264 < \tilde{V}(0{,}05) = 1{,}747$, гипотеза равномерности распределения не отклоняется.

3.4.9. Критерий Фроцини

Ранее мы рассмотрели критерии Фроцини [239] применительно к проверке соответствия эмпирического распределения нормальному (см. раздел 3.2.1.3) и экспоненциальному (см. раздел 3.3.4) законам. Для проверки равномерности ряда $U_i : U_1 \leqslant \ldots \leqslant U_n$ на отрезке $[0, 1]$ статистика критерия Фроцини имеет вид

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| U_i - \frac{i - 0{,}5}{n} \right|.$$

Распределение признается равномерным, если $B_n < B_n(\alpha)$, где $B_n(\alpha)$ — критическое значение статистики Фроцини для проверки равномерности (α — доверительная вероятность), приведенное в табл. 110.

Таблица 110

Критические значения статистики $B_n(\alpha)$ критерия Фроцини для проверки равномерности распределения [239]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
5	0,4964	0,5756	0,7282	11	0,4948	0,5806	0,7486
6	0,4908	0,5700	0,7123	12	0,4987	0,5790	0,7324
7	0,4955	0,5780	0,7428	13	0,4942	0,5815	0,7441
8	0,4933	0,5733	0,7394	14	0,4956	0,5769	0,7417
9	0,4947	0,5764	0,7258	15	0,4961	0,5730	0,7418
10	0,4896	0,5723	0,7310	∞	0,4970	0,5780	0,7440

Задача 157. *Для данных задачи* 149 *проверить гипотезу равномерности распределения случайной величины критерием Фроцини при доверительной вероятности* $\alpha = 0{,}95$.

Вычисляем

$$B_n = \frac{1}{\sqrt{10}} \cdot (|0{,}047 - 0{,}05| + |0{,}05 - 0{,}15| + \ldots + |0{,}91 - 0{,}995|) = 0{,}3298.$$

Из табл. 110 находим $B_n(0{,}95) = 0{,}5723$.

Так как $B_n = 0{,}3298 < B_n(0{,}95) = 0{,}5723$, гипотеза равномерности распределения не отклоняется.

3.4.10. Критерий Гринвуда–Кэсенберри–Миллера

Ранее, при анализе критериев экспоненциальности, мы рассмотрели критерий Гринвуда (см. раздел 3.3.20) [331] со статистикой

$$G = (n+1)\sum_{j=1}^{n+1}(U_j - U_{j-1})^2, \quad U_0 = 0, \quad U_1 \leqslant U_2 \leqslant \ldots \leqslant U_n.$$

В [339] показано, что при $n \to \infty$ распределение G-статистики медленно стремится к стандартному нормальному. Критические значения статистики Гринвуда приведены в табл. 99 (в которой вместо n нужно использовать значения $(n-1)$).

В [350] предложен более мощный критерий проверки равномерности распределения, основанный на статистике

$$Q = \sum_{i=1}^{n+1}(U_i - U_{i-1})^2 + \sum_{i=1}^{n}(U_i - U_{i-1})(U_{i+1} - U_i).$$

Гипотеза равномерности отклоняется, если $Q > Q(\alpha)$, где $Q(\alpha)$ — критическое значение статистики на уровне значимости α, приведенное в табл. 111.

Таблица 111

Критические значения $Q(\alpha)$ статистики Гринвуда–Кэсенберри–Миллера для проверки равномерности распределения [346]

n	Уровень значимости α			n	Уровень значимости α		
	0,01	0,05	0,10		0,01	0,05	0,10
2	0,932	0,859	0,811	32	0,120	0,106	0,101
3	0,831	0,736	0,691	33	0,116	0,103	0,097
4	0,727	0,635	0,586	34	0,112	0,100	0,095
5	0,642	0,551	0,505	35	0,110	0,097	0,092
6	0,573	0,483	0,442	36	0,107	0,095	0,090
7	0,512	0,429	0,393	37	0,103	0,092	0,087
8	0,463	0,387	0,355	38	0,100	0,089	0,085
9	0,423	0,350	0,322	39	0,097	0,087	0,083
10	0,378	0,319	0,294	40	0,095	0,085	0,081
11	0,351	0,294	0,272	41	0,092	0,083	0,079
12	0,318	0,272	0,251	42	0,090	0,081	0,077
13	0,298	0,253	0,234	43	0,088	0,079	0,075
14	0,279	0,237	0,220	44	0,086	0,077	0,073
15	0,259	0,222	0,206	45	0,084	0,075	0,072
16	0,245	0,209	0,195	46	0,082	0,074	0,070
17	0,230	0,197	0,184	47	0,080	0,072	0,069
18	0,218	0,187	0,174	48	0,078	0,070	0,067
19	0,206	0,177	0,166	49	0,077	0,069	0,066
20	0,196	0,168	0,158	50	0,075	0,068	0,065
21	0,187	0,162	0,151	55	0,068	0,061	0,059
22	0,178	0,154	0,144	60	0,062	0,056	0,054
23	0,169	0,147	0,138	65	0,057	0,052	0,050
24	0,163	0,141	0,133	70	0,052	0,048	0,046
25	0,156	0,136	0,128	75	0,048	0,045	0,043
26	0,148	0,131	0,123	80	0,045	0,042	0,040
27	0,144	0,126	0,119	85	0,042	0,039	0,038
28	0,138	0,121	0,114	90	0,040	0,037	0,036
29	0,134	0,117	0,111	95	0,038	0,035	0,034
30	0,130	0,114	0,107	100	0,036	0,033	0,032
31	0,125	0,110	0,104				

Задача 158. *Проверить гипотезу равномерности распределения случайной величины для данных задачи* 149 *критерием Гринвуда–Кэсенберри–Миллера на уровне значимости* $\alpha = 0{,}05$.

Вычисляем

$$\sum_{i=1}^{n+1}(U_i - U_{i-1})^2 = (0{,}047 - 0)^2 + (0{,}05 - 0{,}047)^2 + \ldots + (1 - 0{,}91)^2 = 0{,}127318;$$

$$\sum_{i=1}^{n}(U_i - U_{i-1}) \cdot (U_{i+1} - U_i) = \\ = (0{,}047 - 0) \cdot (0{,}05 - 0{,}047) + \ldots + (0{,}91 - 0{,}72) \cdot (1 - 0{,}91) = 0{,}088259.$$

Тогда $Q = 0{,}127318 + 0{,}088259 = 0{,}215577$.

Из табл. 111 для $\alpha = 0{,}05$ и $n = 10$ находим $Q(0,05) = 0,319$.

Так как $Q = 0{,}215 < Q(0{,}05) = 0{,}319$, гипотеза равномерности распределения не отклоняется.

3.4.11. „Сглаженный“ критерий Неймана–Бартона

Высокую мощность против альтернатив экспоненциального типа имеет критерий равномерности, предложенный в [346]. Критерий, основанный на отношении правдоподобия, строится следующим образом. Для ряда величин $U_1 \leqslant \ldots \leqslant U_n$ строим ряд

$$V_j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \pi_j(U_i),$$

где $\pi_j(U_i)$ — полиномы Лежандра, ортогональные на отрезке $[0, 1]$.

Первые четыре полинома равны

$$\pi_1(y) = 2\sqrt{3}y; \quad \pi_2(y) = \sqrt{5}(6y^2 - 0{,}5);$$
$$\pi_3(y) = \sqrt{7}(20y^3 - 3y^2); \quad \pi_4(y) = 3(70y^4 - 15y^2 + 0{,}375),$$

где $y = U - 0{,}5$.

Статистика Неймана–Бартона имеет вид

$$N_K = \sum_{j=1}^{K} V_j^2.$$

Дэвид [351] показал, что распределение N_K очень хорошо (уже при $n \geqslant 20$) аппроксимируется χ^2-распределением с $f = K$ степенями свободы.

Если $N_K > N_K(\alpha)$, гипотеза равномерности отклоняется на уровне значимости α. Критические значения $N_K(\alpha)$ для $K = 2, 3, 4$ приведены в табл. 112. При $n > 20$ гипотеза равномерности отклоняется, если $N_K > \chi^2_K(\alpha)$, где $\chi^2_K(\alpha)$ — α-квантиль распределения хи-квадрат с $f = K$ степенями свободы.

Задача 159. *Проверить гипотезу равномерности распределения случайной величины для данных задачи* 149 *критерием Неймана–Бартона на уровне значимости* $\alpha = 0{,}05$.

Для ряда

$$U_i\colon 0,047;\ 0,05;\ 0,15;\ 0,18;\ 0,29;\ 0,48;\ 0,52;\ 0,61;\ 0,72;\ 0,91$$

имеем ряд значений $y_i = U_i - 0{,}5$:

$$-0{,}453;\ -0{,}45;\ -0{,}35;\ -0{,}32;\ -0{,}21;\ -0{,}02;\ 0{,}02;\ 0{,}11;\ 0{,}22;\ 0{,}41.$$

Таблица 112

Критические значения $N_K(\alpha)$ критерия равномерности Неймана–Бартона [352, 353]

n	Уровень значимости α			n	Уровень значимости α		
	0,01	0,05	0,10		0,01	0,05	0,10
			N_2				
2	10,012	5,903	4,023	18	9,235	5,918	4,536
3	9,717	5,682	4,013	20	9,234	5,925	4,542
4	9,643	5,566	4,116	25	9,230	5,937	4,554
5	9,517	5,573	4,227	30	9,230	5,947	4,562
6	9,384	5,618	4,316	35	9,226	5,962	4,568
7	9,326	5,640	4,382	40	9,230	5,958	4,573
8	0,276	5,683	4,421	45	9,221	5,961	4,576
9	9,208	5,735	4,453	50	9,223	5,964	4,579
10	9,265	5,774	4,476	60	9,294	5,969	4,584
11	9,173	5,790	4,489	80	9,218	5,974	4,589
12	9,170	5,822	4,486	100	9,220	5,979	4,592
14	9,235	5,897	4,517	∞	9,219	5,991	4,605
16	9,233	5,908	4,527				
			N_3				
2	13,500	7,400	5,590	12	11,560	7,700	6,130
3	12,870	7,480	5,750	14	11,520	7,720	6,140
4	12,450	7,530	5,910	16	11,510	7,730	6,140
5	12,150	7,570	5,990	18	11,500	7,730	6,140
6	11,950	7,600	6,040	20	11,500	7,740	6,150
7	11,810	7,630	6,070	30	11,490	7,750	6,160
8	11,710	7,650	6,100	40	11,490	7,760	6,170
9	11,650	7,670	6,110	50	11,480	7,760	6,180
10	11,600	7,680	6,120	∞	11,350	7,810	6,250
11	11,570	7,690	6,130				
			N_4				
2	16,140	9,520	7,190	12	14,000	9,450	7,650
3	15,800	9,510	7,340	14	13,870	9,440	7,660
4	15,430	9,500	7,460	16	13,780	9,430	7,660
5	15,120	9,490	7,530	18	13,710	9,420	7,670
6	14,860	9,480	7,570	20	13,670	9,420	7,670
7	14,650	9,470	7,600	30	13,580	9,400	7,680
8	14,470	9,470	7,620	40	13,530	9,400	7,680
9	14,320	9,460	7,630	50	13,480	9,400	7,690
10	14,090	9,450	7,650	∞	13,280	9,490	7,780

Вычисляем полиномы для $j = 1$:

$$\pi_1(U_1) = \pi_1(0{,}047) = 2\sqrt{3}\cdot(-0{,}453) = -1{,}569238;$$

$$\pi_1(U_2) = -1{,}55884;\quad \pi_1(U_3) = -1{,}21243;\quad \pi_1(U_4) = -1{,}1085;$$

$$\pi_1(U_5) = -0{,}72746;\quad \pi_1(U_6) = -0{,}062982;\quad \pi_1(U_7) = -0{,}069282;$$

$$\pi(U_8) = 0{,}381051;\quad \pi_1(U_9) = 0{,}762102;\quad \pi_1(U_10) = 1{,}420282.$$

Далее по аналогии вычисляем остальные полиномы $\pi_j(U_i)$. Результаты сведены в таблицу:

i	$\pi_1(U_i)$	$\pi_2(U_i)$	$\pi_3(U_i)$	$\pi_4(U_i)$
1	−1,569238	1,635133	−1,323390	0,129156
2	−1,558840	1,598789	−1,250117	0,623810
3	−1,212430	0,525476	0,509307	−1,236190
4	−1,108510	0,255806	0,806000	−1,280990
5	−0,727460	−0,526370	1,176777	−0,451090
6	−0,069282	−1,112667	0,158322	0,225033
7	−0,069282	−1,112667	−0,158322	0,225033
8	0,381051	−0,955695	−0,802668	0,611246
9	0,762102	−0,468680	−1,182756	−0,561062
10	1,420282	1,137264	0,392680	−0,505402
$\sum$	−3,751607	0,976389	−1,6741667	−2,220456

Далее вычисляем

$$V_1 = \frac{1}{\sqrt{10}} \cdot \sum_{i=1}^{10} \pi_1(U_i) = -1{,}18636230; \quad V_2 = \frac{1}{\sqrt{10}} \cdot \sum_{i=1}^{10} \pi_2(U_i) = 0{,}308761312;$$

$$V_3 = \frac{1}{\sqrt{10}} \cdot \sum_{i=1}^{10} \pi_3(U_i) = -0{,}529417995; \quad V_4 = \frac{1}{\sqrt{10}} \cdot \sum_{i=1}^{10} \pi_4(U_i) = -0{,}70216984;$$

$$N_2 = \sum_{j=1}^{2} V_j^2 = 1{,}502789; \quad N_3 = \sum_{j=1}^{3} V_j^2 = 1{,}783072; \quad N_4 = \sum_{j=1}^{4} V_j^2 = 2{,}2761149.$$

Из табл. 112 для $n = 10$ и $\alpha = 0{,}05$ находим

$$N_2(0{,}05) = 5{,}774, \quad N_3(0{,}05) = 7{,}68 \quad \text{и} \quad N_4(0{,}05) = 9{,}45.$$

Легко видеть, что $N_2 < N_2(\alpha)$, $N_3 < N_3(\alpha)$ и $N_4 < N_4(\alpha)$, что не противоречит гипотезе равномерности распределения.

3.5. Критерии симметрии

Если отсутствуют предпосылки для проверки согласия эмпирического распределения с каким-либо теоретическим, то выявление даже самых общих свойств эмпирического распределения дает некоторую информацию для выбора приемов и методов обработки экспериментального материала.

Одним из таких практически важных свойств распределения является его симметричность относительно центра группирования значений случайной величины.

3.5.1. „Быстрый" критерий Кенуя

Критерий применим для быстрой (но грубой) проверки симметричности в выборках объема $n \geqslant 50$ [121].

Статистика критерия стоится следующим образом. Выборочные значения упорядочиваются по возрастанию: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$, определяются порядковые статистики с номерами $\frac{n}{16}, \frac{15n}{16}, \frac{n}{2}$, т. е. $x_{\left(\frac{n}{16}\right)}, x_{\left(\frac{15n}{16}\right)}, x_{\left(\frac{n}{2}\right)}$.

По ним вычисляются мера асимметрии $A = x_{\left(\frac{15n}{16}\right)} - 2x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{16}\right)}$ и статистика критерия

$$\tilde{A} = \frac{\sqrt{n}}{3s},$$

где $s = \frac{1}{3}\left(x_{\left(\frac{15n}{16}\right)} - x_{\left(\frac{n}{16}\right)}\right)$, или $\tilde{A} = \sqrt{n}\dfrac{x_{\left(\frac{15n}{16}\right)} - 2x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{16}\right)}}{x_{\left(\frac{15n}{16}\right)} - x_{\left(\frac{n}{16}\right)}}$.

При $n \geqslant 50$ распределение статистики может быть аппроксимировано стандартным нормальным распределением. Тогда гипотеза симметрии отклоняется с достоверностью α, если $|\tilde{A}| > u_{\frac{1+\alpha}{2}}$, где u_γ — γ-квантиль стандартного нормального распределения.

Напомним две весьма точные аппроксимации для вычисления нормальных квантилей (см. раздел 1.1.1)

$$u_\gamma = 2{,}0637[\ln(1-\gamma) - 0{,}16]^{0{,}4274} - 1{,}5774$$

с погрешностью менее 0,0008 и $u_\gamma = 4{,}91\left[\gamma^{0,14} - (1-\gamma)^{0,14}\right]$ с погрешностью менее 0,3%.

Задача 160. *В выборке объема* $n = 64$ *имеются порядковые статистики* $x_{\left(\frac{15n}{16}\right)} = x_{(60)} = 127$; $x_{\left(\frac{n}{6}\right)} = x_{(4)} = 17$; $x_{\left(\frac{n}{2}\right)} = x_{(32)} = 44$. *Проверить гипотезу симметричности распределения случайной величины критерием Кенуя при доверительной вероятности* $\alpha = 0{,}95$.

Находим

$$\tilde{A} = \sqrt{64} \cdot \frac{x_{(60)} - 2x_{(32)} + x_{(4)}}{x_{(60)} - x_{(4)}} = 4{,}07.$$

Для $\alpha = 0{,}95$ имеем (применяя аппроксимацию):

$$u_{\frac{1+\alpha}{2}} = u_{0{,}975} = 2{,}0637 \cdot \left[\ln(1 - 0{,}975)^{-1} - 0{,}16\right]^{0{,}4274} - 1{,}5774 = 1{,}96.$$

Так как $\left|\tilde{A}\right| = 4{,}07 > 1{,}96$, гипотеза симметрии отклоняется.

3.5.2. Критерий симметрии Смирнова

Проверяется гипотеза H_0: $F(a) = 1 - F(a)$, то есть гипотеза о том, что предполагаемая функция распределения $F(x)$ симметрична относительно центра a. Для выборки $x_1, x_2, \ldots, x_n$ случайной величины $x > 0$ статистика Смирнова имеет вид [354]

$$\tau_n = \max_{x>0} \left| K^+ - K^- \right|,$$

где K^+ — число значений x, попавших в интервал $(a, a+x)$; K^- — число значений x, попавших в интервал $(a - x, a)$.

Гипотеза симметричности распределения отклоняется с достоверностью α, если $\tau_n \geqslant \tau_n(\alpha)$, где $\tau_n(\alpha)$ — критическое значение статистики, при $n \geqslant 50$ равное [170]

$$\tau_n(\alpha) = \left(n + \frac{2}{3}\right)^{\frac{1}{2}} u_{\frac{1+\alpha}{2}}.$$

Задача 161. *Имеется ряд выборочных данных*

$$x_i: 1,\ 2,\ 3,\ 5,\ 9,\ 11,\ 18,\ 21,\ 29,\ 35.$$

Необходимо проверить гипотезу симметричности распределения случайной величины x относительно $a = 10$ критерием Смирнова при доверительной вероятности $\alpha = 0{,}95$.

Имеем для $x_1 = 1$:
— в интервал $(10, 11)$ попало $K^+ = 1$ значение;
— в интервал $(9, 10)$ попало $K^- = 1$ значение.
Следовательно, при $x_1 = 1$ имеем $K^+ - K^- = 0$.
По аналогии получаем данные, приведенные в таблице:

i	x_i	K^+	K^-	$K^+ - K^-$
1	1	1	1	0
2	2	1	1	0
3	3	1	1	0
4	5	1	2	−1
5	9	2	5	−3
6	11	3	5	−2
7	18	3	5	−2
8	21	4	5	−1
9	29	5	5	0
10	35	5	5	0

Тогда

$$\tau_n \max\left(K^+ - K^-\right) = 3 \quad \text{и} \quad \tau_n(\alpha) = \left(10 + \frac{2}{3}\right)^{\frac{1}{2}} \cdot u_{0,975} = 3{,}266 \cdot 1{,}96 = 6{,}40.$$

Так как $\tau_n = 3 < \tau_n(\alpha) = 6{,}40$, распределение можно считать симметричным (напомним, что критерий следует применять при $n \geqslant 50$, а настоящий пример следует рассматривать как демонстрационный).

3.5.3. Знаковый критерий симметрии

Проверяется гипотеза о симметричности распределения относительно центра. Статистиками критерия являются K^+ — число положительных значений величин $y_i = x_i - a$, и K^- — число отрицательных значений этих величин.

Гипотеза симметричности отклоняется, если

$$\min(K^+, K^-) < K(\alpha),$$

где $K(\alpha)$ — критическое значение, приведенное в табл. 113.

Таблица 113

Критические значения $K(\alpha)$ критерия знаков
(α — уровень значимости) [12]

n	α			n	α			n	α		
	0,01	0,05	0,10		0,01	0,05	0,10		0,01	0,05	0,10
5			0	34	9	10	11	63	20	23	24
6		0	0	35	9	11	12	64	21	23	24
7		0	0	36	9	11	12	65	21	24	25
8	0	0	0	37	10	12	13	66	22	24	25
9	0	1	1	38	10	12	13	67	22	25	26
10	0	1	1	39	11	12	13	68	22	25	26
11	0	1	2	40	11	13	14	69	23	25	27
12	1	2	2	41	11	13	14	70	23	26	27
13	1	2	3	42	12	14	15	71	24	26	28
14	1	2	3	43	12	14	15	72	24	27	28
15	2	3	3	44	13	15	16	73	25	27	28
16	2	3	4	45	13	15	16	74	25	28	29
17	2	4	4	46	13	15	16	75	25	28	29
18	3	4	5	47	14	16	17	76	26	28	30
19	3	4	5	48	14	16	17	77	26	29	30
20	3	5	5	49	15	17	18	78	27	29	31
21	4	5	6	50	15	17	18	79	27	30	31
22	4	5	6	51	15	18	19	80	28	30	32
23	4	6	7	52	16	18	19	81	28	31	32
24	5	6	7	53	16	18	20	82	28	31	33
25	5	7	7	54	17	19	20	83	29	32	33
26	6	7	8	55	17	19	20	84	29	32	33
27	6	7	8	56	17	20	21	85	30	32	34
28	6	8	9	57	18	20	21	86	30	33	34
29	7	8	9	58	18	21	22	87	31	33	35
30	7	9	10	59	19	21	22	88	31	34	35
31	7	9	10	60	19	21	23	89	31	34	36
32	8	9	10	61	20	22	23	90	32	35	36
33	8	10	11	62	20	22	24	91	32	35	37

При $n > 100$ распределение K^+ удовлетворительно аппроксимируется нормальным с $\mathbf{M}(K^+) = \frac{n}{2}$ и $\mathbf{D}(K^+) = \frac{n}{4}$.

Простая формула для критических значений приведена в [12]: $K(\alpha) = \left[\frac{n-1}{2} - c\sqrt{n+1}\right]$, где $[x]$ — ближайшее целое число к x ($c = 1{,}2879$ для $\alpha = 0{,}01$; 0,98 для $\alpha = 0{,}05$ и 0,8224 для $\alpha = 0{,}10$).

Случайная величина $F = \dfrac{K^+}{n - K^- + 1}$ при справедливости нулевой гипотезы симметричности имеет F-распределение Фишера с $f = 2(n - K^+ + 1)$ и $f_2 = 2K^+$ степенями свободы.

Критерий прост в вычислительном отношении, но обладает невысокой мощностью. Поэтому его рекомендуется использовать при $n > 50$ (n — половина объема выборки).

Задача 162. *В условиях задачи* 161 *проверить гипотезу симметричности распределения относительно* $x = 10$ *критерием знаков на уровне значимости* $\alpha = 0{,}05$.

Имеем $K^+ = 5$ и $K^- = 5$, $\min(K^+, K^-) = 5$. Далее

$$K(0{,}05) = \left[\frac{10-1}{2} - 0{,}98 \cdot \sqrt{11}\right] = 1.$$

Так как $\min(K^+, K^-) = 5 > K(0{,}05) = 1$, гипотеза симметричности распределения не отклоняется.

3.5.4. Одновыборочный критерий Вилкоксона

От ряда выборочных величин x_i переходим к ряду величин $y_i = x_i - a$, где a — предполагаемый центр распределения. Значения y_i упорядочим по абсолютной величине: $|y_1| \leqslant |y_2| \leqslant \ldots \leqslant |y_n|$. В полученном ряду каждому значению $|y_i|$ припишем ранг (от 1 до n), равный его порядковому номеру в упорядоченной последовательности.

Обозначим через R_i^+ ранги случайных величин y_i, имеющих положительное значение.

Статистика критерия Вилкоксона задается формулой [355]:

$$T^+ = \sum_{i=1}^{n} R_i^+.$$

Гипотеза симметричности отклоняется, если $T^+ > T^+(\alpha)$, где $T^+(\alpha)$ — критическое значение, приведенное в табл. 114.

Таблица 114

Критические значения $T^+(\alpha)$ одновыборочного критерия Вилкоксона (адаптирована из [24])

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
3	6	6	6	12	55	60	67
4	8	9	10	13	64	69	77
5	12	14	15	14	73	78	88
6	17	18	20	15	82	89	100
7	21	23	27	16	93	99	111
8	27	29	34	17	103	111	124
9	33	36	41	18	115	123	137
10	40	43	49	19	127	136	151
11	47	51	58	20	139	149	166

При $n > 20$ распределение T^+ удовлетворительно аппроксимируется нормальным с параметрами

$$\mathbf{M}(T^+) = \frac{n(n+1)}{4} \quad \text{и} \quad \mathbf{D}(T^+) = \frac{n(n+1)(2n+1)}{24},$$

т. е. критические значения могут быть вычислены по формуле

$$T^{+}(\alpha) = \left[\frac{n(n+1)}{4} + \sqrt{\frac{n(n+1)(2n+1)}{24}}\, u_{\alpha}\right],$$

где $[x]$ — ближайшее целое число к x.

Задача 163. *Для данных задачи* 161 *проверить гипотезу симметричности распределения относительно* $a = 10$ *критерием Вилкоксона при доверительной вероятности* $\alpha = 0{,}95$.

Ряд

$$x_i\colon\ 1,\ 2,\ 3,\ 5,\ 9,\ 11,\ 18,\ 21,\ 29,\ 35$$

преобразуем в ряд

$$y_i = x_i - a\colon\ -9,\ -8,\ -7,\ -5,\ -1,\ 1,\ 8,\ 11,\ 19,\ 25.$$

Ранжированный ряд значений $|y_i|$ имеет вид

1, 1, 5, 7, **8**, 8, 9, **11**, **19**, **25**,

для которого имеем последовательность рангов (для одинаковых значений используются средние ранги):

1,5; 1,5; 3; 4; **5,5**; 5,5; 7; **8**; **9**; **10**.

Отмечая в ряду $|y_i|$ значения $y_i > 0$ (они выделены), приходим к последовательности рангов

$$R_i^{+}\colon\ 1{,}5;\ 5{,}5;\ 8;\ 9;\ 10.$$

Тогда $T^{+} = \sum\limits_{i=1}^{n} R_i^{+} = 1{,}5 + 5{,}5 + 8 + 9 + 10 = 34$.

Так как $T^{+} = 34 < T^{+}(\alpha) = 43$, гипотеза симметричности не отклоняется.

Для нормального приближения ($u_{0,95} = 1{,}645$) имеем

$$T^{+}(0{,}95) = \left[\frac{10\cdot 11}{4} + \sqrt{\frac{10\cdot 11\cdot 21}{24}}\cdot 1{,}645\right] = 43,$$

что совпадает с табличным результатом.

3.5.5. Критерий Антилла–Керстинга–Цуккини

Критерий, рассмотренный в [356], использует статистику, подобную статистике Вилкоксона (см. раздел 3.5.4).

Пусть $\tilde{x}$ — эмпирическая медиана ряда $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$. Образуем ряд величин $y_i = |x_i - \tilde{x}|$, и пусть R_i — ранг величины y_i в упорядоченном по возрастанию ряду значений y_i.

Обозначим

$$\operatorname{sign}(z) = \begin{cases} 1, & z > 0; \\ 0, & z = 0; \\ -1, & z < 0. \end{cases}$$

Тогда статистика критерия имеет вид

$$T_1(\gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} c_{\gamma}\left(\frac{R_i}{2(n+1)}\right) \operatorname{sign}(x_i - \tilde{x}),$$

$$\text{где}\quad c_{\gamma} = \min\left(\alpha, \frac{\gamma}{2}\right) \quad \text{при} \quad 0 \leqslant a,\ \gamma \leqslant \frac{1}{2}.$$

Гипотеза симметричности отклоняется, если $T_1(\gamma) \geqslant T_{\text{кр}}(\lambda)$.

При $\gamma = 0$ и $n \geqslant 7$ статистика

$$T_1(0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} c_0\left(\frac{R_i}{2(n+1)}\right) \operatorname{sign}(x_i - \tilde{x})$$

распределена асимптотически нормально с $\mu = 0$ и дисперсией $\sigma^2 = 0{,}03156 \exp\left(-\frac{a}{n}\right)$, где $a = 5{,}66$ для четных n и $a = 9{,}75$ для нечетных n. При $\gamma = \frac{1}{6}$ для $T\left(\frac{1}{6}\right)$ также применима нормальная аппроксимация при $n \geqslant 7$ с дисперсией $\sigma^2 = 0{,}01736 \exp\left(-\frac{a}{n}\right)$, где $a = 5{,}77$ для четных n и $a = 11{,}11$ для нечетных n.

Таким образом, для рассмотренных статистик критические значения могут быть вычислены по формулам (α — доверительная вероятность):

— для четных n

$$T_{\text{кр}}(0) = 0{,}1776513 \exp\left(-\frac{2{,}83}{n}\right) u_\alpha; \quad T_{\text{кр}}\left(\frac{1}{6}\right) = 0{,}131757 \exp\left(-\frac{2{,}885}{n}\right) u_\alpha;$$

— для нечетных n

$$T_{\text{кр}}(0) = 0{,}177613 \exp\left(-\frac{4{,}875}{n}\right) U_\alpha; \quad T_{\text{кр}}\left(\frac{1}{6}\right) = 0{,}131757 \exp\left(-\frac{5{,}555}{n}\right) u_\alpha,$$

где u_α — α-квантиль стандартного нормального распределения. Критерии этой группы обладают большей мощностью, чем критерий Вилкоксона (см. раздел 5.4).

Задача 164. *Для данных задачи 161 проверить гипотезу симметричности распределения критериями $T_1(0)$ и $T_1(1/6)$ Антилла–Керстинга–Цуккини при $\alpha = 0{,}95$.*

Для ряда

$$x_i\colon 1,\ 2,\ 3,\ 5,\ 9,\ 11,\ 18,\ 21,\ 29,\ 35$$

медиана равна $\tilde{x} = \frac{x_5 + x_6}{2} = 10$. Получаем ряд значений

$$y_i = |x_i - \tilde{x}|: 9,\ 8,\ 7,\ 4,\ 1,\ 1,\ 8,\ 11,\ 19,\ 25.$$

Ранжированному ряду

$$y_i\colon 1,\ 1,\ 4,\ 7,\ 8,\ 8,\ 9,\ 11,\ 19,\ 25$$

соответствует последовательность рангов

$$R_i\colon 1{,}5;\ 1{,}5;\ 3;\ 4;\ 5{,}5;\ 5{,}5;\ 7;\ 8;\ 9;\ 10.$$

Поэтому последовательность рангов R_i, соответствующая последовательности y_i, имеет вид

$$7;\ 5,5;\ 4;\ 3;\ 1,5;\ 1,5;\ 5,5;\ 8;\ 9;\ 10.$$

Для этой последовательности ряд $\operatorname{sign}(x_i - \tilde{x})$ будет иметь вид

$$-1,\ -1,\ -1,\ -1,\ -1,\ +1,\ +1,\ +1,\ +1,\ +1.$$

Вычисляем статистики критерия:

$$T_1\,(10) = \frac{1}{\sqrt{n}} \cdot \sum_{i=1}^{n} \min\left[\frac{R_i}{2\,(n+1)}, \frac{1}{2}\right] \cdot \operatorname{sign}\,(x_i - \tilde{x}) = \frac{1}{\sqrt{10}} \times$$
$$\times \left[\min\left(\frac{7}{2 \cdot 11}, \frac{1}{2}\right) \cdot (-1) + \min\left(\frac{5{,}5}{2 \cdot 11}, \frac{1}{2}\right) \cdot (-1) + \ldots + \min\left(\frac{10}{2 \cdot 11}, \frac{1}{2}\right) \cdot 1\right] = 0{,}1868;$$

$$T_1\left(\frac{1}{6}\right) = \frac{1}{\sqrt{n}} \cdot \sum_{i=1}^{n} \min\left[\frac{R_i}{2(n+1)}; \frac{1}{2} - \frac{1}{6}\right] \cdot \operatorname{sign}(x_i - \tilde{x}) = \frac{1}{\sqrt{10}} \times$$

$$\times \left[\min\left(\frac{7}{2 \cdot 11}, \frac{1}{3}\right) \cdot (-1) + \min\left(\frac{5{,}5}{2 \cdot 11}, \frac{1}{3}\right) \cdot (-1) + \ldots + \min\left(\frac{10}{2 \cdot 11}, \frac{1}{3}\right) \cdot 1\right] = 0{,}11498.$$

Для $\alpha = 0{,}95$ имеем $u_\alpha = 1{,}645$, и критические значения статистик равны:

$$T_{\text{кр}}(0) = 0{,}17765 \cdot \exp\left(-\frac{2{,}83}{10}\right) \cdot 1{,}645 = 0{,}220;$$

$$T_{\text{кр}}\left(\frac{1}{6}\right) = 0{,}131757 \cdot \exp\left(-\frac{2{,}885}{10}\right) \cdot 1{,}645 = 0{,}1624.$$

Так как $T_1(0) = 0{,}1868 < T_{\text{кр}}(0) = 0{,}220$ и $T_1\left(\frac{1}{6}\right) = 0{,}115 < T_{\text{кр}}\left(\frac{1}{6}\right) = 0{,}162$, гипотеза симметричности распределения случайных величин не отклоняется.

3.5.6. Критерий Бхатачарья–Гаствирта–Райта (модифицированный критерий Вилкоксона)

Критерий является обобщением критерия Вилкоксона (см. раздел 3.5.4) при проверке гипотезы относительно неизвестного центра распределения [357].

Статистика критерия строится следующим образом. Пусть имеется выборка $x_1, \ldots, x_{2n+1}$ и определены вероятности p и $q = 1 - p$ $(0 < p < 1)$. Обозначим выборочные квантили $x_{[n+1]}$, $x_{[nq+1]}$ и $x_{[2n+1-nq]}$ ($x_{[\beta]}$ — $[\beta]$-я порядковая статистика в упорядоченном по возрастанию ряду x_i; $[\ldots]$ — целая часть числа, содержащегося в скобках).

В [357] предложены две статистики. Первая статистика является модификацией критерия Вилкоксона (см. раздел 5.4) в форме, рассмотренной авторами работы [358].

Обозначим

$$y_{(i)} = x_{[nq+1]} - x_{[nq+1-i]}; \quad z_{(i)} = x_{[2n+1-nq+i]} - x_{[2n+1-nq]};$$

$$\chi(u) = \begin{cases} 1, & u < 0; \\ \frac{1}{2}, & u = 0, \quad i = 1, \ldots, nq; \\ 0, & u > 0. \end{cases}$$

Тогда статистика критерия записывается в форме

$$T = \frac{1}{(nq)^2} \sum_{i=1}^{nq} \sum_{j=1}^{nq} \chi(y_{(i)} - z_{(j)}).$$

Статистика основана на сравнении $\frac{1}{2}q$-х верхней и нижней частей данных. Вторая статистика построена в ранговой форме. Пусть

$$y_{(i)} = x_{n+1} - x_{n+1-i}; \quad z_{(i)} = x_{n+1-i} - x_{n+1}.$$

Объединим y' и z' в одну выборку, обозначим через $R_1, \ldots, R_n$ и $R_{n+1}, \ldots, R_{2n}$ ранги y'-ов и z'-ов в объединенной выборке. Тогда статистика критерия записывается в форме

$$S = \frac{1}{(2n+1)^{\frac{3}{2}}} \left\{ \sum_{i=1}^{n} (R_i - 2np)^+ - \sum_{i=n+1}^{2n} (R_i - 2np)^+ \right\}, \quad \text{где} \quad x^+ = \max(x, 0).$$

При $n \geqslant 10$ можно считать, что распределение статистик T и S нормально со средними $\mathbf{M}(T) = \frac{1}{2}$ и $\mathbf{M}(S) = 0$ и дисперсиями [357]:

— для $q = 0{,}25$: $\mathbf{D}(T) = 1$ и $\mathbf{D}(S) = 0{,}006$;

— для $q = 0{,}5$: $\mathbf{D}(T) = 0{,}754$ и $\mathbf{D}(S) = 0{,}9394$.

Случай $q = 1$ соответствует обычному критерию Вилкоксона с центром, оцениваемым выборочной медианой. Критерии рассмотренного типа обладают большей мощностью, чем критерий Вилкоксона.

Гипотеза симметричности распределения отклоняется с доверительной вероятностью α, если:

— при $q = 0{,}25$

$$|T| > T_{\text{кр}} = 0{,}5 + 1{,}187u_{\frac{1+\alpha}{2}}; \quad |S| > S_{\text{кр}} = 0{,}0774u_{\frac{1+\alpha}{2}};$$

— при $q = 0{,}5$

$$|T| > T_{\text{кр}} = 0{,}5 + 0{,}8683u_{\frac{1+\alpha}{2}}; \quad |S| > S_{\text{кр}} = 0{,}1985u_{\frac{1+\alpha}{2}},$$

где u_γ — γ-квантиль стандартного нормального распределения.

Задача 165. *Проверить гипотезу симметричности распределения критериями T и S Бхатачарья–Гаствирта–Райта при $q = 0{,}5$ и $\alpha = 0{,}95$ для выборки данных*

$$x_i\colon\ 1,\ 2,\ 3,\ 5,\ 9,\ 1,\ 18,\ 21,\ 29,\ 35,\ 45,\ 49,\ 51.$$

Имеем $2n + 1 = 13$, $n = 6$ и

$$y_{(i)} = x_{[nq+1]} - x_{[nq+1-i]} = x_{[6\cdot 0{,}5+1]} - x_{[6\cdot 0{,}5+1-i]} = x_{[4]} - x_{[4-i]};$$
$$z_{(i)} = x_{[2n+1-nq+i]} - x_{[2n+1-nq]} = x_{[10+i]} - x_{[10]}.$$

Вычисляем

$$T = \frac{1}{(6\cdot 0{,}5)^2}\cdot\sum_{i=1}^{3}\sum_{j=1}^{3}\chi(y_{(i)} - z_{(j)}) =$$
$$= \frac{1}{9}\cdot\{\chi(y_{(1)} - z_{(1)}) + \chi(y_{(1)} - z_{(2)}) + \chi(y_{(1)} - z_{(3)}) + \chi(y_{(2)} - z_{(1)}) + \ldots + \chi(y_{(3)} - z_{(3)})\} =$$
$$= \frac{1}{9}\cdot\{\chi(x_4 - x_3 - x_{13} + x_{10}) + \chi(x_4 - x_3 - x_{12} + x_{10}) + \ldots + \chi(x_4 - x_1 - x_{13} + x_{10})\} =$$
$$= \frac{1}{9}\cdot\{\chi(5 - 3 - 49 + 35) + \chi(5 - 3 - 51 + 35) + \ldots + \chi(5 - 1 - 51 + 35)\} =$$
$$= \frac{1}{9}\cdot\{\chi(-12) + \chi(-14) + \ldots + \chi(-12)\} = 1.$$

Тогда при $u_{\frac{1+0{,}95}{2}} = u_{0{,}975} = 1{,}96$ имеем $T_{\text{кр}} = 0{,}5 + 0{,}8683\cdot 1{,}96 = 2{,}20$.

Так как $T = 1 < T_{\text{кр}} = 2{,}20$, гипотеза симметричности распределения не отклоняется.

Рассмотрим теперь S-критерий. Вычисления, необходимые для подсчета значения критерия, приведены в таблице ($n = 6$, $x_{n+1} = x_7 = 18$):

i	x_{n+1-i}	x_{n+1+i}	$y'_{(i)}$	$z'_{(i)}$	$y'(z')$	$R_i(y'_i)$	$R_i(z'_i)$
1	11	21	7	3	3(1)		1
2	9	29	9	11	7(2)	2	
3	5	35	13	17	9(3)	3	
4	3	45	15	27	11(4)		4
5	2	49	16	31	13(5)	5	
6	1	51	17	33	15(6)	6	

Окончание

i	x_{n+1-i}	x_{n+1+i}	$y'_{(i)}$	$z'_{(i)}$	$y'(z')$	$R_i(y'_i)$	$R_i(z'_i)$
7					16(7)	7	
8					17(8,5)	8,5	
9					17(8,5)		8,5
10					27(10)		10
11					31(11)		11
12					33(12)		12

Статистика критерия равна

$$s = \frac{1}{(2\cdot 6+1)^{\frac{3}{2}}}\cdot\left\{\max(2-2\cdot 6\cdot 0{,}5;0)+\max(3-2\cdot 6\cdot 0{,}5;0)+\ldots+\max(12-6;0)\right\} =$$
$$= 0{,}2133\cdot(0+0+\ldots+2{,}5-0\ldots-6) = -0{,}2982.$$

Далее имеем $S_{\text{кр}} = 0{,}1985\cdot 1{,}96 = 0{,}389$. Так как $|S| = 0{,}298 < S_{\text{кр}} = 0{,}389$, гипотеза симметричности не отклоняется.

3.5.7. Критерий Финча

Критерий рассмотрен в [359]. Основан на интервалах порядковых статистик, обладает высокой чувствительностью к отклонению распределения от симметричности.

Для выборки $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ введем обозначения

$$R_i = x_{n-k+i} - x_{n-i-k+i};\quad L = x_{k+2-i} - x_{k+1-i};\quad V_i = \frac{R_i - L_i}{R_i + L_i},$$

где $i = 1, \ldots, k$; $k = \left[\frac{n}{2}\right]$.

Статистика критерия имеет вид $V = |\nu_{ii}\sum V_i\omega_i|$, где

$$\omega_i = c\sum_{j=k-(i-1)}^{k+(i-1)}\frac{1}{j};$$

$$c = \left(\sum_{i=1}^{k}\omega_i^2\right)^{-1};\quad \nu = 2d^2 + \left(1-d^2\right)\left(am^2+bm+\frac{1}{3}\right)^{-1};$$

$$d = 0{,}278 + \frac{1{,}8}{n};\quad a = 0{,}41 + \frac{23{,}4}{n^2};\quad m = 1{,}47 - \frac{8}{n};\quad b = -2ma.$$

При $n \geqslant 10$ применима аппроксимация критических значений V-статистики (α — доверительная вероятность)

$$V_{\text{кр}} = \nu\sum_{i=1}^{k}\omega_i + 0{,}5773\nu u_\alpha,$$

где u_α — α-квантиль стандартного нормального распределения.

Если

$$|V| \geqslant \nu\left(\sum_{i=1}^{k}\omega_i + 0{,}5773u_\alpha\right),$$

то гипотеза симметричности распределения отклоняется с доверительной вероятностью α.

Задача 166. *Для данных задачи* 165 *проверить гипотезу симметричности критерием Финча при доверительной вероятности* $\alpha = 0{,}95$.

Имеем ряд

$$x_i\colon\ 1,\ 2,\ 3,\ 5,\ 9,\ 11,\ 21,\ 29,\ 35,\ 45,\ 49,\ 51\ (n = 13).$$

Находим

$$d = 0{,}278 + \frac{1{,}6}{13} = 0{,}41646; \quad m = 1{,}47 - \frac{8}{13} = 0{,}8546; \quad a = 0{,}41 + \frac{23{,}4}{13^2} = 0{,}54846;$$

$$b = -2 \cdot 0{,}8546 \cdot 0{,}54846 = -0{,}93743; \quad \nu = 3d^2 + \left(1 - d^2\right) \cdot \left(am^2 + bm + \frac{1}{3}\right)^{-1} = -11{,}774.$$

Далее вычисляем

$$k = \left[\frac{n}{2}\right] = 6;$$

$$\omega_j = c \cdot \sum_{j=7-i}^{5+i} \frac{1}{j}; \quad \omega_1 = c \cdot 0{,}16667; \quad \omega_2 = c \cdot \sum_{j=5}^{7} \frac{1}{j} = c \cdot 0{,}5095;$$

$$\omega_3 = c \cdot 0{,}8845; \quad \omega_4 = c \cdot 1{,}32897; \quad \omega_5 = c \cdot 1{,}928968; \quad \omega_6 = c \cdot 3{,}01987.$$

Из равенства $\sum \omega_i^2 = c^2 \cdot 15{,}6764 = 1$ находим $c = 0{,}252567$.

Далее

$$V_1 = \frac{R_1 - L_1}{R_1 + L_1} = \frac{x_{n-k+1} - x_{n-1-k+1} - x_{k+2-1} + x_{k+1-1}}{x_{n-k+1} - x_{n-1-k+1} + x_{k+2-1} - x_{k+1-1}} =$$

$$= \frac{x_{13-6+1} - x_{13-1-6+1} - x_{6+2-1} + x_{6+1-1}}{x_{13-6+1} - x_{13-1-6+1} + x_{6+2-1} - x_{6+1-1}} = \frac{x_8 - 2x_7 + x_6}{x_8 - x_6} = \frac{21 - 2 \cdot 18 + 11}{21 - 11} = 0{,}4.$$

С учетом полученного значения $c = 0{,}25256$ имеем

$$\omega_1 = \frac{0{,}25256}{6} = 0{,}04209; \quad \omega_2 = 0{,}12860; \quad \omega_3 = 0{,}22339; \quad \omega_4 = 0{,}33564;$$

$$\omega_5 = 0{,}48718; \quad \omega_6 = 0{,}762698.$$

Вычисляем статистику критерия

$$|V| = |-11{,}774 \cdot (0{,}04209 \cdot (-0{,}4) + 0128603 \cdot 0{,}6 + \ldots + 0{,}762698 \cdot 0{,}3333)| = 10{,}3058.$$

Далее находим критическое значение ($u_{0{,}95} = 1{,}645$)

$$V_{\text{кр}} = \left|\nu \cdot \left(\sum_{i=1}^{k} \omega_i^2 + 0{,}5773 \cdot 1{,}645\right)\right| = 34{,}4876.$$

Так как $V = 10{,}3058 < V_{\text{кр}} = 34{,}4876$, гипотеза симметричности распределения не отклоняется.

3.5.8. Критерий Бооса

Критерий рассмотрен в [360], позволяет проверить симметричность распределения относительно его центра, в качестве которого используется оценка Ходжеса–Лемана

$$\tilde{\tilde{x}} = \operatorname{med}\left\{\frac{x_i + x_j}{2}\right\}, \quad 1 \leqslant i \leqslant j \leqslant n,$$

т. е. медиана всех возможных $\dfrac{n \cdot (n+1)}{2}$ попарных средних Уолша. Статистика

критерия имеет вид

$$B = n\left\{-1 + \frac{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i + x_j - 2\tilde{\tilde{x}}\right|}{\sum_{i<j}^{n}\left|x_i - x_j\right|}\right\}.$$

Для $n \geqslant 10$ критические точки $B(\alpha)$ приведены в табл. 115.

Таблица 115

Критические значения *B*-критерия Бооса

(α — доверительная вероятность) [360]

α	0,90	0,95	0,975	0,99
$B(\alpha)$	0,712	0,899	1,093	1,358

Если $B > B(\alpha)$, то гипотеза симметричности отклоняется. Критерий обладает высокой эффективностью и менее чувствителен, чем другие, к выбросам на хвостах распределения.

Задача 167. *Для данных задачи* 165 *проверить гипотезу симметричности распределения критерием Бооса при* $\alpha = 0{,}95$.

Имеем ряд

$$x_i\colon\ 1,\ 2,\ 3,\ 5,\ 9,\ 11,\ 18,\ 21,\ 29,\ 35,\ 45,\ 49,\ 51.$$

Сначала вычислим оценку Ходжеса–Лемана. Для нахождения медианы расположим все значения $\frac{x_i + x_j}{2}$ (их будет $\frac{n(n+1)}{2} = \frac{13 \cdot 14}{2} = 91$) в ряд по возрастанию и найдем медиану $\tilde{\tilde{x}} = x_{\frac{n+1}{2}} = x_{\frac{91+1}{2}} = x_{46} = 20$ (т. е. 46-ю порядковую статистику).

Дальнейшие вычисления сведены в таблицу, в которой приняты обозначения $\tilde{x}^{*}_{i,j} = \left|x_i + x_j - 2\tilde{\tilde{x}}\right|$; $\tilde{x}^{**}_{i,j} = \left|x_i - x_j\right|$:

i	j	$\tilde{x}^{*}_{ij}$	$\tilde{x}^{**}_{ij}$	i	j	$\tilde{x}^{*}_{ij}$	$\tilde{x}^{**}_{ij}$	i	j	$\tilde{x}^{*}_{ij}$	$\tilde{x}^{**}_{ij}$
1	1	40		2	7	21	16	3	13	13	58
1	2	38	1	2	8	18	19	4	1	35	
1	3	37	2	2	9	10	27	4	2	34	
1	4	35	4	2	10	4	33	4	3	33	
1	5	31	8	2	11	5	43	4	4	31	
1	6	29	10	2	12	10	47	4	5	27	4
1	7	22	17	2	13	12	49	4	6	25	6
1	8	19	20	3	1	37		4	7	18	13
1	9	11	28	3	2	36		4	8	15	16
1	10	5	34	3	3	35		4	9	7	24
1	11	5	44	3	4	33	2	4	10	1	30
1	12	9	48	3	5	29	6	4	11	9	40
1	13	11	50	3	6	27	8	4	12	13	44
2	1	40		3	7	20	15	4	13	15	46
2	2	37		3	8	17	18	5	1	31	
2	3	36	1	3	9	9	26	5	2	30	
2	4	34	3	3	10	3	323	5	3	29	
2	5	30	7	3	11	7	42	5	4	27	
2	6	28	9	3	12	11	46	5	5	23	

Окончание

i	j	$\tilde{x}^*_{ij}$	$\tilde{x}^{**}_{ij}$	i	j	$\tilde{x}^*_{ij}$	$\tilde{x}^{**}_{ij}$	i	j	$\tilde{x}^*_{ij}$	$\tilde{x}^{**}_{ij}$
5	6	21	2	8	5	11		11	4	9	
5	7	14	9	8	6	9		11	5	13	
5	8	11	12	8	7	2		11	6	15	
5	9	3	20	8	8	1		11	7	22	
5	10	3	26	8	9	9	8	11	8	15	
5	11	13	36	8	10	15	14	11	9	33	
5	12	17	40	8	11	15	24	11	10	39	
5	13	19	42	8	12	29	28	11	11	49	
6	1	29		8	13	31	30	11	12	53	4
6	2	28		9	1	11		11	13	55	6
6	3	27		9	2	10		12	1	9	
6	4	25		9	3	9		12	2	10	
6	5	21		9	4	7		12	3	11	
6	6	19		9	5	3		12	4	13	
6	7	12	7	9	6	1		12	5	17	
6	8	9	10	9	7	6		12	6	19	
6	9	1	10	9	8	9		12	7	26	
6	10	5	24	9	9	17		12	8	29	
6	11	15	34	9	10	23		12	9	37	
6	12	19	38	9	11	33	33	12	10	43	
6	13	21	40	9	12	37	20	12	11	53	
7	1	22		9	13	39	22	12	12	57	
7	2	21		10	1	5		12	13	59	2
7	3	20		10	2	4		13	1	11	
7	4	18		10	3	3		13	2	12	
7	5	14		10	4	1		13	3	13	
7	6	12		10	5	3		13	4	15	
7	7	5		10	6	5		13	5	19	
7	8	2	3	10	7	12		13	6	21	
7	9	6	11	10	8	15		13	7	28	
7	10	12	17	10	9	23		13	8	31	
7	11	22	27	10	10	29		13	9	39	
7	12	26	31	10	11	39	10	13	10	45	
7	13	28	33	10	12	43	43	13	11	55	
8	1	19		10	13	45	16	13	12	59	
8	2	18		11	1	5		13	13	61	
8	3	17		11	2	5					
8	4	15		11	3	7					

Суммируя столбцы таблицы, получаем

$$\sum_{i=1}^{n}\sum_{i=1}^{n}|x_i + x_j - 2\tilde{\tilde{x}}| = 3532 \quad \text{и} \quad \sum_{i<j}^{n}|x_i - x_j| = 1688.$$

Окончательно вычисляем

$$B = 13 \cdot \left\{-1 + \frac{0{,}5 \cdot 3532}{1688}\right\} = 0{,}6007.$$

Из табл. 15 для $\alpha = 0{,}95$ находим $B(0{,}95) = 0{,}899$.

Так как $B = 0{,}6007 < B(0{,}95) = 0{,}899$, гипотеза симметричности распределения не отклоняется.

3.5.9. Критерий Гупты

Критерий рассмотрен в [314, 361] и обладает достаточной мощностью против обширного класса альтернатив.

Последовательность процедур вычисления критерия включает в себя следующие шаги. Для выборки объема $x_1 \leqslant \ldots \leqslant x_n$ находим медиану

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}}, & n = 1, 3, \ldots, 2k-1; \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right), & n = 2, 4, \ldots, 2k. \end{cases}$$

Вводим переменную

$$\delta_{ij} = \begin{cases} 1, & \text{если } x_i + x_j > 2\tilde{x}; \\ 0, & \text{если } x_i + x_j \leqslant 2\tilde{x} \end{cases}$$

и находим

$$A_1 = \sum_{j=2}^{n} \sum_{i=1}^{j-1} \delta_{ij}; \quad A_2 = \frac{1}{2}\left[\frac{n-2}{2}\right]\left[\frac{n}{2}\right],$$

где $[y]$ — наибольшее целое, меньшее или равное y.

Далее вычисляем

$$c = \frac{2}{n(n+1)}(A_1 - A_2)\,.$$

Вводим переменную

$$a_i = \begin{cases} 1, & \text{если } \left(\tilde{x} - n^{-\frac{1}{5}}\right) \leqslant x_i \leqslant \left(\tilde{x} + n^{-\frac{1}{5}}\right); \\ 0, & \text{если неравенство не выполняется.} \end{cases}$$

Находим

$$D = \max\left\{1, \sum_{i=1}^{n} a_i\right\}; \quad A_3 = \left(2n^{\frac{4}{5}}\right)D; \quad t = \frac{u_{1-\frac{\alpha}{2}}}{\Delta\tilde{x}\sqrt{3n}},$$

где u_y — y-квантиль стандартного нормального распределения; $\Delta\tilde{x}$ — симметричный двусторонний доверительный интервал для $\tilde{x}$.

Построение симметричного двустороннего доверительного интервала для $\tilde{x}$ с доверительной вероятностью $(1-\alpha)$ включает следующие шаги. Находим величину

$$k(\alpha) = \left[\frac{n(n+1)}{4} + \sqrt{\frac{n(n+1)(2n+1)}{24}} u_{\frac{\alpha}{2}}\right],$$

где $[\ldots]$ — ближайшее целое к числу в скобках.

Далее составляем $N = \dfrac{n(n+1)}{2}$ величин $z_{ij} = \dfrac{x_i + x_j}{2} (i < j)$ и ранжируем их по возрастанию.

Искомый интервал равен

$$\Delta\tilde{x} = z_{[k]} - z_{[l]}, \text{где} \quad k = N + 1 - k(\alpha)\,; \quad l = k(\alpha)\,.$$

Графический метод определения величины $\Delta\tilde{x}$ изложен в [362].

Продолжая алгоритм вычисления критерия Гупты, находим

$$V = \frac{1}{12} + \frac{1}{4}\left(1 - \frac{2t}{A_3}\right)^2; \quad J = \left(\frac{V}{n}\right)^{-\frac{1}{2}}\left(c - \frac{1}{4}\right).$$

Если $|J| \geqslant u_{1-\frac{\alpha}{2}}$, то гипотеза симметричности отклоняется с доверительной вероятностью $1 - \alpha$.

Задача 168. *Для выборки*

$$x_i\colon 1,\ 2,\ 3,\ 5,\ 9,\ 11,\ 18,\ 21,\ 29,\ 35$$

проверить гипотезу симметричности распределения критерием Гупты при доверительной вероятности 0,90.

Находим оценку медианы (n — четное): $\tilde{x} = \dfrac{x_5 + x_6}{2} = \dfrac{9 + 11}{2} = 10.$

Для вычисления A_1 составим таблицу величин δ_{ij} при различных сочетаниях i и j:

j	i								
	1	2	3	4	5	6	7	8	9
2	0								
3	0	0							
4	0	0	0						
5	0	0	0	0					
6	0	0	0	0	0				
7	0	0	1	1	1	1			
8	1	1	1	1	1	1	1		
9	1	1	1	1	1	1	1	1	
10	1	1	1	1	1	1	1	1	1

Находим $A_1 = 28$ (число единиц в таблице) и $A_2 = \dfrac{\left[\dfrac{10-2}{2}\right] \cdot \left[\dfrac{10}{2}\right]}{2} = \dfrac{4 \cdot 5}{2} = 10.$

Далее вычисляем

$$c = \frac{2}{10 \cdot 9} \cdot (28 - 10) = 0{,}4; \quad \tilde{x} - n^{-\frac{1}{5}} = 10 - 10^{-\frac{1}{5}} = 9{,}369; \quad \tilde{x} + n^{-\frac{1}{5}} = 10{,}631.$$

Имеем $a_1 = a_2 = \ldots = a_{10} = 0$; $D = \max\{1{,}0\} = 1$; $A_3 = 2 \cdot 10^{\frac{4}{5}} \cdot 1 = 12{,}619.$

Из условия задачи $1 - \alpha = 0{,}90$ ($\alpha = 0{,}10$), следовательно, $u_{\frac{\alpha}{2}} = u_{0{,}05} = -1{,}645.$

Найдем $N = \dfrac{n(n+1)}{2} = \dfrac{10 \cdot 11}{2} = 55$ значений полусумм $z_{ij} = \dfrac{x_i + x_j}{2}$ $(i < j)$, которые представлены в таблице.

Имеем ряд (равные ранги усредняются):

i	j	z_{ij}	Ранг	i	j	z_{ij}	Ранг	i	j	z_{ij}	Ранг
1	1	1	1	2	2	2	3,5	3	4	4	9
1	2	1,5	2	2	3	2,5	5	3	5	6	13,5
1	3	2	3,5	2	4	3,5	8	3	6	7	16,5
1	4	3	6,5	2	5	5,5	12	3	7	10,5	23
1	5	5	10,5	2	6	5,5	12	3	8	12	28
1	6	6	13,5	2	7	10	21,5	3	9	16	36,5
1	7	9,5	20	2	8	11,5	26,5	3	10	19	41,5
1	8	11	24,5	2	9	15,5	35	4	4	5	10,5
1	9	15	33,5	2	10	13,5	30,5	4	5	7	16,5
1	10	18	39,5	3	3	3	6,5	4	6	8	18

Окончание

i	j	z_{ij}	Ранг	i	j	z_{ij}	Ранг	i	j	z_{ij}	Ранг
4	7	11,5	26,5	5	10	22	47	7	10	26,5	52
4	8	13	29	6	6	11	24,5	8	8	21	46
4	9	17	38	6	7	14,5	32	8	9	25	51
4	10	19,5	43,5	6	8	16	36,5	8	10	23	48,5
5	5	9	19	6	9	20	45	9	9	29	53
5	6	10	21,5	6	10	23	48,5	9	10	32	54
5	7	13,5	30,5	7	7	18	39,5	10	10	35	55
5	8	15	33,5	7	8	19,5	43,5				
5	9	19	40,5	7	9	23,5	50				

Находим $k(\alpha) = \left[\dfrac{10\cdot 11}{2} + \sqrt{\dfrac{10\cdot 11\cdot 21}{24}}\cdot(-1{,}645)\right] = 11.$

Следовательно,

$$k = N + 1 - k(\alpha) = 55 + 1 - 11 = 45; \quad l = k(\alpha) = 11; \quad \Delta\tilde{x} = z_{45} - z_{11} = 20 - 5 = 15.$$

Далее вычисляем

$$t = \frac{u_{1-\frac{\alpha}{2}}}{\Delta\tilde{x}\cdot\sqrt{3n}} = \frac{u_{0,95}}{\Delta\tilde{x}\cdot\sqrt{3n}} = \frac{1{,}645}{15\cdot\sqrt{30}} = 0{,}02; \quad V = \frac{1}{12} + \frac{1}{4}\cdot\left(1 - \frac{2\cdot 0{,}02}{12{,}619}\right)^2 = 0{,}3317;$$

$$J = \left(\frac{V}{n}\right)^{-\frac{1}{2}}\cdot(c - 0{,}25) = \left(\frac{0{,}3317}{10}\right)^{-\frac{1}{2}}\cdot(0{,}4 - 0{,}25) = 0{,}823.$$

Так как $J = 0{,}823 < u_{0,95} = 1{,}645$, гипотеза симметричности распределения не отклоняется.

3.5.10. Критерий Фрезера

Пусть $a_n(i) = M(x_i)$ — математическое ожидание модуля i-й порядковой статистики в выборке объема n из стандартного нормального распределения. По аналогии с критерием знаков (см. раздел 5.3), от переменной x_i переходим к переменной $y_i = x_i - a$ (a — предполагаемый центр распределения). Упорядочиваем y_i по абсолютной величине: $|y_1| \leqslant |y_2| \leqslant \ldots \leqslant |y_n|$. Пусть R_i — ранг в этом ряду положительных значений $y_i = x_i - a > 0$. Фрезер [363, 364] предложил для проверки симметричности распределения случайных величин статистику

$$S = \sum_{i=1}^{n} a_n(R_i).$$

Для практических расчетов применима следующая аппроксимация с помощью квантилей стандартного нормального распределения:

$$a_n(R_i) = u_{\beta_i}, \quad \text{где} \quad \beta_i = \frac{1}{2}\left(\frac{R_i - \frac{3}{8}}{n + \frac{1}{4}} + 1\right).$$

Напомним ранее приведенные аппроксимации (см. раздел 5.1)

$$u_{\alpha_i} = 2{,}0637\left[\ln(1-\alpha_i)^{-1} - 0{,}16\right]^{0,4247} - 1{,}5774$$

($\alpha_i > 0{,}5$; при $\alpha_i < 0{,}5$ следует вместо α_i использовать $1-\alpha_i$ и поменять знак у u_{α_i});

$$u_{\alpha_i} = 4{,}91\left[\alpha_i^{0,14} - (1-\alpha_i)^{0,14}\right].$$

При $n \geqslant 10$ распределение S может быть аппроксимировано нормальным распределением [365] со средним $\mathbf{M}(S)$ и дисперсией $\mathbf{D}(S)$, где

$$\mathbf{M}(S) = \frac{n}{\sqrt{2\pi}} = 0{,}399n; \quad \mathbf{D}(S) = \frac{1}{4}\sum_{i=1}^{n}[a_n(R_i)]^2 = \frac{1}{4}\sum_{i=1}^{n}U_{\beta_i}^2.$$

Следовательно, случайная величина $\dfrac{S - 0{,}399n}{0{,}5\sqrt{\sum\limits_{i=1}^{n} u_{\beta_i}^2}}$ будет иметь стандартное нормальное распределение. Поэтому критическое значение S-статистики равно

$$S(\alpha) = 0{,}399 \cdot n + 0{,}5 \cdot \sqrt{u_{\beta_i}^2} \cdot u_\alpha.$$

Если $\sum\limits_{i=1}^{n} u_{\beta_i}^2 \geqslant 0{,}399n + 0{,}5\sqrt{u_{\beta_i}^2}\,u_\alpha$, то с доверительной вероятностью α гипотеза симметричности распределения отклоняется.

Задача 169. *Для данных задачи* 168 *проверить гипотезу симметрии распределения критерием Фрезера при* $\alpha = 0{,}95$ *относительно медианы, равной* 10.

Имеем ряд

$$y_i = x_i - \tilde{x}\colon\ -9;\ -8;\ -7;\ -5;\ -1;\ 1;\ 7;\ 11;\ 19;\ 25.$$

Строим ранжированный ряд

$$|y_i| = |x_i - \tilde{x}|\colon\ 1;\ 1;\ 5;\ 7;\ 7;\ 8;\ 9;\ 11;\ 19;\ 25.$$

Ему соответствует последовательность рангов

$$1,5;\ 1,5;\ 3;\ 4,5;\ 4,5;\ 6;\ 7;\ 8;\ 9;\ 10.$$

Легко убедиться, что значения $y_i = x_i - \tilde{x} > 0$ (это 1, 7, 11, 19 и 25) имеют в общем ряду ранги $R_i = 1,5;\ 4,5;\ 8;\ 9;\ 10$. Рангам R_i соответствует ряд значений:

$$\beta_1 = \frac{1}{2}\cdot\left(\frac{R_1 - \frac{3}{8}}{n + \frac{}{4}} + 1\right) = \frac{1}{2}\cdot\left(\frac{1{,}5 - 0{,}375}{10{,}25} + 1\right) = 0{,}55488;$$

$$\beta_2 = \frac{1}{2}\cdot\left(\frac{4{,}5 - 0{,}375}{10{,}25} + 1\right) = 0{,}70122; \quad \beta_3 = \frac{1}{2}\cdot\left(\frac{8 - 0{,}375}{20{,}25} + 1\right) = 0{,}87195;$$

$$\beta_4 = \frac{1}{2}\cdot\left(\frac{9 - 0{,}375}{10{,}25} + 1\right) 0{,}92073; \qquad \beta_5 = \frac{1}{2}\cdot\left(\frac{10 - 0{,}375}{10{,}25} + 1\right) = 0{,}969512.$$

Соответствующие им квантили u_{β_i} равны

$$u_{\beta_1} = 4{,}91\cdot\left[\beta_1^{0{,}14} - (1-\beta_1)^{0{,}14}\right] = 4{,}91\cdot\left[0{,}55488^{0{,}14} - (1 - 0{,}55488)^{0{,}14}\right] = 0{,}13738;$$

$$u_{\beta_2} = 0{,}52597; \quad u_{\beta_3} = 1{,}13446; \quad u_{\beta_4} = 1{,}4104; \quad u_{\beta_5} = 1{,}87673.$$

Вычисляем

$$S = \sum_i u_{\beta_i} = 5{,}08494; \quad \sum_i u_{\beta_i}^2 = 7{,}09386.$$

Для $\alpha = 0{,}95$ и $u_{0{,}95} = 1{,}645$ имеем

$$S(0{,}95) = 8{,}399 \cdot 10 + 0{,}5 \cdot (7{,}09386)^{\frac{1}{2}} \cdot 1{,}645 = 6{,}1807.$$

Так как $S = 5{,}085 < S(0{,}95) = 6{,}18$, гипотеза симметричности распределения совокупности случайных величин не отклоняется.

3.6. Подбор кривых распределения вероятностей по экспериментальным данным

Часто инженер или исследователь не имеет достаточных оснований для выбора того или иного закона распределения вероятностей. Его попытка использовать критерии нормальности, экспоненциальности или равномерности распределения случайных величин потерпели неудачу. Что же ему делать? Очевидно, смириться с тем, что он не сможет применить хорошо известные математические модели для описания своих экспериментальных данных, и попытаться найти все-таки модель, которой отвечают его результаты.

Другими словами, ему необходимо подобрать по экспериментальным данным распределение, которое удовлетворительно описывало бы имеющийся экспериментальный материал.

3.6.1. Кривые распределения Джонсона

Джонсон [366] предложил для аппроксимации эмпирических распределений использовать кривые, получающиеся с помощью преобразований плотности нормального распределения.

Пусть x — случайная величина, для которой подбирается распределение Джонсона. В общем случае преобразование Джонсона имеет вид

$$z = \gamma + \eta f(x; \varepsilon, \lambda), \quad \eta > 0, \quad -\infty < \gamma < \infty, \quad \lambda > 0, \quad -\infty < \varepsilon < \infty,$$

где γ, η, ε и λ — параметры распределения Джонсона; $f(\ldots)$ — произвольная функция; z — случайная величина, имеющая стандартное нормальное распределение.

Джонсон [13] предложил три формы функционального преобразования:

$$f_1(x; \varepsilon, \lambda) = \ln\left(\frac{x-\varepsilon}{\lambda}\right), \qquad x \geqslant \varepsilon;$$

$$f_2(x; \varepsilon, \lambda) = \ln\left(\frac{x-\varepsilon}{\lambda+\varepsilon-x}\right), \qquad \varepsilon \leqslant x \leqslant \varepsilon+\lambda;$$

$$f_3(x; \varepsilon, \lambda) = \operatorname{Arsh}\left(\frac{x-\varepsilon}{\lambda}\right), \quad -\infty < x < \infty.$$

Им соответствуют семейства кривых S_L, S_B и S_U Джонсона.

Подбор кривых распределения Джонсона по экспериментальным данным проводится в два этапа. Сначала определяется, какое из трех семейств распределений приемлемо. Затем находятся оценки параметров выбранного семейства распределений (S_L, S_B или S_U).

Чтобы решить, какое распределение из трех семейств распределения Джонсона следует использовать для описания полученных экспериментальных данных, вычисляем эмпирические оценки третьего (α_3) и четвертого (α_4) моментов

$$\hat{\alpha}_3 = \frac{1}{ns^3}\sum_{i=1}^{n}(x_i-\bar{x})^3; \quad \hat{\alpha}_4 = \frac{1}{ns^4}\sum_{i=1}^{n}(x_i-\bar{x})^4, \quad \text{где} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

Далее рекомендуется следующая приближенная процедура выбора типа распределения:

если $\hat{\alpha}_4 > 3(1+0{,}641\hat{\alpha}_3^2)$ — выбирается семейство S_U;

если $\hat{\alpha}_4 \approx 3(1+0{,}641\hat{\alpha}_3^2)$ — выбирается семейство S_L;

если $\hat{\alpha}_4 < 3(1+0{,}641\hat{\alpha}_3^2)$ — выбирается семейство S_B.

При $\alpha_4 < 1 + \alpha_3$ кривые Джонсона неприменимы. Более точный метод выбора вида распределения Джонсона по значениям α_3, α_4 и коэффициента вариации $v = \frac{\mu}{s}$ приведен в [367]. Обширный анализ вопросов, связанных с семейством распределений Джонсона, содержится в [368].

3.6.1.1. Семейство распределений S_L Джонсона

Плотность распределения Джонсона семейства S_L имеет вид

$$f_1(x) = \frac{\eta}{\sqrt{2\pi}\,(x-\varepsilon)} \exp\left\{-\frac{1}{2}\eta^2\left[\frac{\tilde{\gamma}}{\eta} + \ln(x-\varepsilon)\right]^2\right\},$$

где $\eta > 0$; $-\infty < \tilde{\gamma} = \gamma - \eta \ln \lambda < \infty$; $-\infty < \varepsilon < \eta$; $-\infty < \gamma < \infty$; $\lambda > 0$ — параметры распределения и $x \geqslant \varepsilon$.

Распределение, соответствующее кривой $f_1(x)$, является логарифмически нормальным (см. раздел 1.1.3) с параметрами $\eta = \frac{1}{\sigma}$, $\gamma = -\frac{\mu}{\sigma}$ и параметром положения ε, определяющим начало координат.

Рассмотрим два случая: значение ε известно (положим $\varepsilon = 0$) и значение параметра неизвестно. В первом случае переходим от переменной x к переменной $\hat{x} = x - \varepsilon$. Тогда оценки параметров η и $\hat{\gamma}$ имеют вид $\hat{\eta} = \frac{1}{s}$ и $\hat{\hat{\gamma}} = -\frac{m}{s}$, где

$$m = \frac{1}{n}\sum_{i=1}^{n} \ln x_i, \qquad s^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} \ln^2 x_i - \left(\sum_{i=1}^{n} \ln x_i\right)^2\right].$$

Если значение ε неизвестно, то оценки параметров находятся по формулам

$$\hat{\eta} = 1{,}645\left[\ln\left(\frac{x_{[0{,}95n]} - x_{[0{,}5n]}}{x_{[0{,}5n]} - x_{[0{,}05n]}}\right)\right]^{-1}; \quad \hat{\hat{\gamma}} = \hat{\eta}\ln\left[\frac{1 - \exp\left(-\frac{1{,}645}{\hat{\eta}}\right)}{x_{[0{,}5n]} - x_{[0{,}05n]}}\right];$$

$$\hat{\varepsilon} = x_{[0{,}5n]} - \exp\left(-\frac{\hat{\hat{\gamma}}}{\hat{\eta}}\right),$$

где $x_{[\beta]}$ — порядковая статистика эмпирического распределения с номером β.

Рекомендуется сначала найти квантили распределения случайной величины $\ln x$ с помощью таблиц (или формул) стандартного нормального распределения, а затем потенцированием перейти к квантилям исходной случайной величины x.

Задача 170. *В результате испытаний* $n = 100$ *изделий получены следующие значения случайной переменной* x (*параметра изделия*):

1,0	5,1	6,3	7,5	8,7	9,6	10,8	11,9	13,5	15,3
2,0	5,2	6,5	7,5	8,7	9,7	10,9	11,9	13,9	15,9
2,5	5,3	6,6	7,9	8,8	9,8	10,9	12,0	14,0	16,0
3,0	5,3	6,7	7,9	8,8	9,9	11,0	12,3	14,1	16,5
3,5	5,5	6,8	8,0	8,9	10,1	11,0	12,4	14,2	17,0
4,2	5,6	7,0	8,1	9,0	10,2	11,2	12,5	14,3	17,5
4,6	5,8	7,0	8,1	9,2	10,2	11,3	12,7	14,4	18,0
4,9	6,0	7,1	8,3	9,4	10,2	11,5	12,7	14,5	21,2
5,0	6,1	7,1	8,4	9,5	10,5	11,7	13,0	14,8	22,3
5,0	6,1	7,3	8,5	9,5	10,6	11,8	13,2	15,2	26,1

Необходимо выяснить возможность аппроксимации распределения переменной x кривой из семейства Джонсона и оценить параметры этого распределения. Вычислить вероятность того, что параметр изделия $x \leqslant 15$. Найти значение параметра изделия, вероятность превышения которого не более 0,05.

Сначала вычислим необходимые исходные данные s^2, α_3 и α_4.
Имеем

$$\bar{x} = 9{,}971; \quad s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = 19{,}075; \quad S = 4{,}36749;$$

$$\alpha_3 = \frac{1}{n \cdot s^3} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^3; \quad \alpha_4 = \frac{1}{n \cdot s^4} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^4 = 4{,}18945.$$

Находим, что $\alpha_4 = 4{,}189 \approx 3 \cdot \left(1 + 0{,}641 \cdot \alpha_3^2\right) = 4{,}143$. Поэтому для описания полученных данных применимо семейство S_L Джонсона. Вычисляем параметры распределения

$$m = \frac{1}{n} \cdot \sum_{i=1}^{n} \ln x_i = 2{,}1915; \quad s = \left[\frac{n \cdot \sum\limits_{i=1}^{n} \ln^2 x_i - \left(\sum\limits_{i=1}^{n} \ln x_i \right)^2}{n(n-1)} \right] = 0{,}0505798.$$

В нашем случае порядковые статистики равны:

$$x_{[0{,}95 \cdot n]} = x_{0{,}95} = 17{,}0; \quad x_{[0{,}05 \cdot n]} = x_5 = 3{,}5; \quad x_{[0{,}5 \cdot n]} = x_{50} = 9{,}5.$$

Далее

$$\hat{\eta} = 1{,}645 \cdot \left[\ln\left(\frac{17 - 9{,}5}{9{,}5 - 3{,}5} \right) \right]^{-1} = 7{,}3719;$$

$$\hat{\gamma} = 7{,}3719 \cdot \ln\left[\frac{1 - \exp\left(1 - \dfrac{1{,}645}{7{,}3719}\right)}{9{,}5 - 3{,}5} \right] = -25{,}0732; \quad \hat{\varepsilon} = 9{,}5 - \exp\left(\frac{25{,}0732}{7{,}3719} \right) = -20{,}5.$$

Таким образом, плотность распределения случайной величины может быть описана функцией

$$f(x) = \frac{7{,}3719}{\sqrt{2\pi} \cdot (x + 20{,}5)} \cdot \exp\left\{ -\frac{1}{2} \cdot 7{,}3719^2 \cdot \left[-\frac{25{,}0732}{7{,}3719} + \ln(x + 20{,}5) \right]^2 \right\} =$$
$$= 2{,}941 \cdot (x + 20{,}5)^{-1} \cdot \exp\left\{ -29{,}172 \cdot [-3{,}401 + \ln(x + 20{,}5)]^2 \right\}.$$

Случайная величина $z = \tilde{\gamma} + \eta \cdot \ln(x - \varepsilon)$ имеет стандартное нормальное распределение, что позволяет выразить квантили u_α^x распределения случайной величины x через квантили u_α стандартного нормального распределения

$$u_\alpha^x = \exp\left[\frac{1}{\hat{\eta}} \cdot \left(u_\alpha - \hat{\tilde{\gamma}} \right) \right] + \hat{\varepsilon}.$$

Этим соотношением можно воспользоваться для ответа на дальнейшие вопросы, поставленные в задаче. Найдем сначала вероятность того, что $x \leqslant 15$. Задача сводится к следующему — по заданной квантили u_α^x найти u_α и вычислить вероятность, соответствующую этой квантили. Имеем

$$15 = \exp\left[\frac{1}{7{,}3719} \cdot (u_\alpha + 25{,}0732) \right] - 20{,}5.$$

Отсюда $u_\alpha = 1{,}241$, и искомая величина α ищется из условия $\alpha = F(u_\alpha)$, где $F(y)$ — функция стандартного нормального распределения.

Для вычисления $F(y)$ воспользуемся аппроксимацией (см. раздел 1.1.1)

$$F(y) = 1 - 0{,}852 \cdot \exp\left\{-\left(\frac{y + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\}$$

и получим

$$\alpha = F(u_\alpha) = 1 - 0{,}852 \cdot \exp\left\{-\left(\frac{1{,}241 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\} = 0{,}893.$$

Следовательно, вероятность того, что $x \leqslant 15$, равна 0,893.

Вторая задача — вычисление значения u_α^x, для которого $P(x > u_\alpha^x) = 0{,}05$, эквивалентна поиску квантили u_α^x при $\alpha = 1 - 0{,}05 = 0{,}95$. Имеем $u_{0{,}95} = 1{,}645$ и

$$u_\alpha^x = \exp\left[\frac{1}{7{,}3719} \cdot (1{,}645 + 25{,}0732)\right] - 20{,}5 = 17.$$

Следовательно, вероятность того, что $x \geqslant 17$, не более 0,05.

3.6.1.2. Семейство распределений S_B Джонсона

Уравнение для плотности распределения кривых этого семейства имеет вид

$$f(x) = \frac{\eta}{\sqrt{2\pi}}\frac{\lambda}{(x-\varepsilon)(\lambda - x + \varepsilon)} \exp\left\{-\frac{1}{2}\left[\gamma + \eta \ln\left(\frac{x-\varepsilon}{\lambda - x + \varepsilon}\right)\right]^2\right\}, \quad \varepsilon \leqslant x \leqslant \varepsilon + \lambda,$$

где $\eta > 0$, $-\infty < \gamma < \infty$, $\lambda > 0$, $-\infty < \varepsilon < \infty$ — параметры распределения.

Случайная величина x, имеющая распределение из семейства S_B Джонсона, ограничена пределами ε и $\varepsilon + \lambda$.

Возможны три случая:

— оба предела ε и $\varepsilon+\lambda$ известны (например, если случайной величиной является доля дефектных изделий в партии, заключенная, очевидно, между 0 и 1);

— известен один предел (например, если x — долговечность изделия, то $\varepsilon = 0$);

— оба предела неизвестны.

Если оба предела известны (известны параметры ε и λ), то оценки параметров η и γ находятся по формулам [13, 366]

$$\hat{\eta} = \frac{u_{\alpha''} - u\alpha'}{\ln\left[\dfrac{(x_{\alpha''} - \varepsilon)(\varepsilon + \lambda - x_{\alpha'})}{(x_{\alpha'} - \alpha)(\varepsilon + \lambda - x_{\alpha''})}\right]}; \quad \hat{\gamma} = u_{\alpha''} - \eta \ln\left(\frac{x_{\alpha''} - \varepsilon}{\varepsilon + \lambda - x_{\alpha''}}\right),$$

где u_α — α-квантиль стандартного нормального распределения; x_α — эмпирическая квантиль ($[\alpha(n+1)]$-й упорядоченный по возрастанию член выборки $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$).

Рекомендуется выбирать $\alpha' = \alpha$, $\alpha'' = 1 - \alpha$ и уровень $\alpha = 0{,}05$.

Квантили аппроксимирующего распределения u_α^x могут быть выражены через квантили стандартного нормального распределения u_α следующим образом:

$$u_\alpha^x = \frac{\varepsilon + (\lambda + \varepsilon)\exp\left(\dfrac{u_\alpha - \gamma}{\eta}\right)}{1 + \exp\left(\dfrac{u_\alpha - \gamma}{\eta}\right)}.$$

Если известно только одно крайнее значение ε, то в дополнение к оценкам γ и η

необходимо найти оценку λ по формуле

$$\hat{\lambda} = (\tilde{x} - \varepsilon)\frac{(\tilde{x} - \varepsilon)(x_\alpha - \varepsilon) + (\tilde{x} - \varepsilon)(x_{1-\alpha} - \varepsilon) - 2(x_\alpha - \varepsilon(x_{1-\alpha} - \varepsilon))}{(\tilde{x} - \varepsilon)^2 - (x_\alpha - \varepsilon)(x_{1-\alpha} - \varepsilon)},$$

где $\tilde{x}$ — выборочная медиана ряда.

Параметры η и γ находятся по формулам для случая, когда оба предела известны (вместо параметра λ используется его эмпирическая оценка $\hat{\lambda}$). Если пределы ε и $\varepsilon + \lambda$ неизвестны, то для нахождения оценок параметров необходимо решить численными методами систему четырех нелинейных уравнений

$$u_{\alpha_i} = \gamma + \eta \ln\left(\frac{x_{\alpha_i} - \varepsilon}{\varepsilon + \lambda - x_{\alpha_i}}\right); \quad i = 1, 2, 3, 4,$$

где u_{α_i} и x_{α_i} — α-квантили соответствующих величин.

Однако такая ситуация практически очень редка и здесь не рассматривается.

Задача 171. *Наблюдение за временем изготовления $n = 50$ изделий позволило получить следующие данные (мин):*

0,70	0,93	1,09	1,21	1,41
0,72	0,95	1,10	1,23	1,46
0,75	0,95	1,10	1,23	1,54
0,81	0,96	1,11	1,25	1,55
0,83	0,98	1,12	1,29	1,61
0,84	0,98	1,13	1,32	1,61
0,87	1,02	1,14	1,34	1,62
0,89	1,04	1,15	1,37	1,62
0,90	1,06	1,16	1,38	1,63
0,91	1,08	1,17	1,40	1,63

Необходимо установить возможность аппроксимации распределения времени изготовления изделий кривой из семейства S_B Джонсона. Найти параметры распределения, вычислить вероятность того, что искомая величина будет находиться в интервале $0{,}92 \leqslant x \leqslant 1{,}28$.

Вычисляем

$$\bar{x} = 1{,}1628;\ s^2 = 0{,}07086;\ s = 0{,}26620;\ \sum_{i=1}^{n}(x_i - \bar{x})^3 = 0{,}257706;\ \sum_{i=1}^{n}(x_i - \bar{x})^4 = 0{,}51103;$$

$$\hat{\alpha}_3 = \frac{1}{ns^3} \cdot \sum_{i=1}^{n}(x_i - \bar{x})^3 = 0{,}273231; \quad \hat{\alpha}_4 = \frac{1}{ns^4} \cdot \sum_{i=1}^{n}(x_i - \bar{x})^4 = 2{,}03551.$$

Имеем $\hat{\alpha}_4 = 2{,}03551 < 3 \cdot (1 + 0{,}641 \cdot \hat{\alpha}_3^2) = 3{,}143$, что соответствует области семейства S_B. В нашем случае по смыслу задачи $\varepsilon = 0$ и необходимо найти оценки параметров η, γ и λ. Выбирая $\alpha = 0{,}05$, имеем $u_{0,95} = 1{,}645$ и $u_{0,05} = -1{,}645$. Далее находим статистики с порядковыми номерами $[0{,}05 \cdot 51] = 2$ $(x_2 = 0{,}72)$; $[(1-\alpha)\cdot(n+1)] = [0{,}95 \cdot 51] = 48$ $(x_{48} = 1{,}62)$.

Медиана выборки равна $\tilde{x} = \dfrac{1}{2} \cdot (x_{25} + x_{26}) = \dfrac{1{,}12 + 1{,}13}{2} = 1{,}125.$

Тогда для $\varepsilon = 0$ находим

$$\hat{\lambda} = 1{,}125 \cdot \frac{1{,}125 \cdot 0{,}72 + 1{,}125 \cdot 1{,}62 - 2 \cdot 0{,}72 \cdot 1{,}62}{1{,}125^2 - 0{,}72 \cdot 1{,}62} = 3{,}398;$$

$$\hat{\eta} = \frac{1{,}645 + 1{,}645}{\ln\left[\dfrac{1{,}62 \cdot (3{,}398 - 0{,}72)}{0{,}72 \cdot (3{,}398 - 1{,}62)}\right]} = 2{,}695; \quad \hat{\gamma} = 1{,}645 - 2{,}695 \cdot \ln\left(\frac{1{,}62}{3.398 - 1{,}62}\right) = 1{,}896.$$

Таким образом, плотность аппроксимирующего распределения есть

$$f(x) = \frac{2{,}695}{\sqrt{n}} \cdot \exp\left\{-\frac{1}{2} \cdot \left[1{,}896 + 2{,}695 \cdot \ln\left(\frac{x}{3{,}398 - \bar{x}}\right)\right]^2\right\} =$$

$$= \frac{3{,}653}{x \cdot (3{,}398 - x)} \cdot \exp\left\{-\left[1{,}341 + 1{,}906 \cdot \ln\left(\frac{x}{3{,}398 - x}\right)\right]^2\right\}.$$

Квантиль распределения u_α^x равна

$$u_\alpha^x = \frac{\hat{\pi} \cdot \exp\left(\dfrac{u_\alpha - \hat{\gamma}}{\hat{\eta}}\right)}{1 + \exp\left(\dfrac{u_\alpha - \hat{\gamma}}{\hat{\eta}}\right)} = \frac{3{,}398 \cdot \exp(0{,}371 \cdot u_\alpha - 0{,}703)}{1 + \exp(0{,}371 \cdot u_\alpha - 0{,}703)}.$$

Вероятность того, что $0{,}92 \leqslant x \leqslant 1{,}28$, равна по определению

$$\mathbf{P}(0{,}92 \leqslant x \leqslant 1{,}28) = \mathbf{P}(x \leqslant 1{,}28) - \mathbf{P}(x \leqslant 0{,}92).$$

Задача сводится к тому, чтобы найти вероятности $\beta' = \mathbf{P}(x \leqslant 1{,}28)$ и $\beta'' = \mathbf{P}(x \leqslant 0{,}92)$, соответствующие квантилям $u_{\beta'}^x = 1{,}28$ и $u_{\beta''}^{x=0{,}92}$.

Алгоритм решения этой задачи включает в себя следующие этапы:

— по заданным $u_{\beta'}^x$ и $u_{\beta''}^x$ находятся соответствующие им значения $u_{\beta'}$ и $u_{\beta''}$;

— по найденным значениям $u_{\beta'}$ и $u_{\beta''}$ вычисляются значения функции стандартного нормального распределения $F(u_{\beta'})$ и $F(u_{\beta''})$;

— вычисляется искомая вероятность $p = \beta' - \beta''$.

Теперь, пользуясь аппроксимацией для $F(y)$ (см. предыдущую задачу), вычисляем

$$\beta' = F(U_\beta') = F(0{,}5374) = 1 - 0{,}852 \cdot \exp\left\{-\left(\frac{0{,}5374 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\} = 0{,}704;$$

$$\beta'' = F(U_\beta'') = F(-0{,}7758) = 1 - 0{,}852 \cdot \exp\left\{-\left(\frac{-0{,}7758 + 1{,}5774}{2{,}0637}\right)^{2{,}34}\right\} = 0{,}236.$$

Окончательно получаем $p = \beta' - \beta'' = 0{,}704 - 0{,}236 = 0{,}468$.

Следовательно, вероятность того, что $0{,}92 \leqslant x \leqslant 2{,}28$, равна 0,468.

3.6.1.3. Семейство распределений S_U Джонсона

Уравнение кривых распределения семейства S_U Джонсона имеет вид

$$f(x) = \frac{\eta}{\sqrt{2\pi}} \frac{1}{\sqrt{(x-\varepsilon)^2 + \lambda^2}} \exp\left\{-\frac{1}{2}\left(\gamma + \eta \ln\left\{\frac{x-\varepsilon}{\lambda} + \left[\left(\frac{x-\varepsilon}{\lambda}\right)^2\right]^{\frac{1}{2}}\right\}\right)^2\right\},$$

$$-\infty < x < \infty, \quad \eta > 0, \quad -\infty < \gamma < \infty, \quad \lambda > 0, \quad -\infty < \varepsilon < \infty.$$

Случайная величина, имеющая распределение S_U Джонсона, теоретически неограничена, и в общем случае все четыре параметра γ, η, λ и ε неизвестны и должны быть оценены по выборке. Для оценки параметров η и γ следует использовать табл. 116, в которой α_3 и α_4 — соответственно третий и четвертый моменты распределения вероятностей.

Затем вычисляются оценки

$$\hat{\lambda} = s\left\{\frac{1}{2}(\omega - 1)\left[\omega \operatorname{ch}\left(\frac{2\hat{\gamma}}{\hat{\eta}}\right) + 1\right]\right\}^{-\frac{1}{2}}; \quad \varepsilon = \bar{x} + \hat{\lambda}\sqrt{\omega}\operatorname{sh}\left(\frac{\hat{\gamma}}{\hat{\eta}}\right), \quad \text{где} \quad \omega = \exp\left(\frac{1}{\hat{\eta}^2}\right).$$

Таблица 116

Оценки параметров γ и η распределения S_U Джонсона [369]

α_4	α_3							
	0,05		0,10		0,15		0,20	
	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η
3,2	0,3479	4,671	0,7373	4,787	1,2280	5,004	1,9490	5,369
3,3	0,2328	3,886	0,4834	3,927	0,7747	4,036	1,1430	4,208
3,4	0,1760	3,396	0,3620	3,435	0,5699	3,503	0,8166	3,607
3,5	0,1421	3,081	0,2905	3,108	0,4528	3,156	0,6384	3,227
3,6	0,1196	2,852	0,2435	2,872	0,3776	2,908	0,5260	2,960
3,7	0,1035	2,676	0,2102	2,692	0,3238	2,719	0,4487	2,760
3,8	0,0914	2,535	0,1853	2,548	0,2845	2,571	0,3921	2,604
3,9	0,0820	2,420	0,1661	2,431	0,2542	2,450	0,3490	2,477
4,0	0,0745	2,324	0,1507	2,333	0,2302	2,349	0,3150	2,372
4,1	0,0684	2,242	0,1381	2,250	0,2106	2,264	0,2875	2,283
4,2	0,0633	2,171	0,1276	2,178	0,1943	2,190	0,2647	2,207
4,3	0,0589	2,109	0,1188	2,115	0,1806	2,126	0,2456	2,141
4,4	0,0552	2,054	0,1112	2,060	0,1689	2,069	0,2294	2,082
4,5	0,0519	2,005	0,1046	2,010	0,1588	2,018	0,2153	2,030
4,6	0,0491	1,961	0,0989	1,966	0,1499	1,973	0,2031	1,984
4,7	0,0466	1,921	0,0938	1,925	0,1421	1,932	0,1923	1,942
4,8	0,0444	1,885	0,0893	1,889	0,1352	1,895	0,1828	1,904
4,9	0,0424	1,852	0,0852	1,855	0,1290	1,861	0,1743	1,869
5,0	0,0406	1,822	0,0816	1,825	0,1234	1,830	0,1666	1,837
5,1	0,0390	1,793	0,0783	1,796	0,1184	1,801	0,1597	1,808
5,2	0,0374	1,767	0,0752	1,770	0,1138	1,775	0,1534	1,781
5,3	0,0361	1,743	0,0725	1,746	0,1096	1,750	0,1477	1,756
5,4	0,0348	1,721	0,0700	1,723	0,1057	1,727	0,1424	1,732
5,5	0,0337	1,699	0,0676	1,702	0,1022	1,705	0,1376	1,711
5,6	0,0326	1,680	0,0655	1,682	0,0989	1,685	0,1331	1,690
5,7	0,0316	1,661	0,0635	1,663	0,0958	1,666	0,1290	1,671
5,8	0,0307	1,643	0,0616	1,645	0,0930	1,648	0,1251	1,653
5,9	0,0298	1,627	0,0599	1,628	0,0904	1,631	0,1215	1,636
6,0	0,0290	1,611	0,0583	1,613	0,0879	1,615	0,1182	1,619
6,1	0,0283	1,596	0,0568	1,598	0,0856	1,606	0,1151	1,604
6,2	0,0276	1,582	0,0553	1,583	0,0835	1,586	0,1121	1,590
6,3	0,0269	1,568	0,0540	1,570	0,0814	1,572	0,1094	1,576
6,4	0,0263	1,556	0,0527	1,557	0,0795	1,559	0,1067	1,563
6,5	0,0257	1,543	0,0515	1,545	0,0777	1,555	0,1043	1,550
6,6	0,0251	1,532	0,0504	1,533	0,0760	1,554	0,1020	1,538
6,7	0,0246	1,520	0,0493	1,522	0,0743	1,524	0,0998	1,527
6,8	0,0241	1,510	0,0483	1,511	0,0728	1,513	0,0977	1,516
6,9	0,0236	1,499	0,0473	1,501	0,0713	1,502	0,0957	1,505
7,0	0,0232	1,490	0,0464	1,491	0,0699	1,492	0,0938	1,495
7,1	0,0277	1,480	0,0455	1,481	0,0686	1,448	0,0920	1,485
7,2	0,0223	1,471	0,0447	1,472	0,0673	1,474	0,0903	1,476
7,3	0,0219	1,462	0,0439	1,463	0,0661	1,465	0,0887	1,467
7,4	0,0215	1,454	0,0431	1,455	0,0650	1,456	0,0871	1,458
7,5	0,0212	1,445	0,0424	1,446	0,0639	1,448	0,0856	1,450
7,6	0,0208	1,438	0,0417	1,438	0,0628	1,440	0,0842	1,442
7,7	0,0205	1,430	0,0410	1,431	0,0618	1,432	0,0828	1,434
7,8	0,0202	1,415	0,0404	1,423	0,0608	1,425	0,0815	1,427
7,9	0,0198	1,415	0,0398	1,416	0,0599	1,418	0,0802	1,419
8,0	0,0195	1,408	0,0392	1,409	0,0590	1,411	0,0790	1,412
8,0	0,0195	1,408	0,0392	1,409	0,0590	1,411	0,0790	1,412
8,2	0,0190	1,395	0,0380	1,396	0,0572	1,397	0,0767	1,399
8,4	0,0185	1,383	0,0370	1,383	0,0557	1,385	0,0745	1,386
8,6	0,0180	1,371	0,0360	1,372	0,0542	1,373	0,0725	1,374
8,8	0,0175	1,360	0,0351	1,361	0,0528	1,362	0,0707	1,363
9,0	0,0171	1,349	0,0342	1,350	0,0515	1,351	0,0689	1,352

Продолжение таблицы 116

α_4	α_3 0,25 $-\gamma$	0,25 η	0,30 $-\gamma$	0,30 η	0,35 $-\gamma$	0,35 η	0,40 $-\gamma$	0,40 η
3,2	3,1890	5,992	6,3890	7,204				
3,3	1,6560	4,469	2,4770	4,875				
3,4	1,1330	3,759	1,5690	3,979	2,2360	4,300	3,4840	4,813
3,5	0,8620	3,328	1,1480	3,467	1,5460	3,663	2,1460	3,943
3,6	0,6997	3,033	0,9115	3,132	1,1870	3,266	1,5650	3,448
3,7	0,5907	2,816	0,7586	2,890	0,9681	2,989	1,2380	3,120
3,8	0,5125	2,648	0,6515	2,707	0,8197	2,783	1,0280	2,882
3,9	0,4536	2,513	0,5723	2,561	0,7127	2,623	0,8814	2,701
4,0	0,4076	2,402	0,5113	2,442	0,6304	2,492	0,7733	2,557
4,1	0,3707	2,309	0,4629	2,343	0,5673	2,385	0,6902	2,439
4,2	0,3404	2,229	0,4234	2,258	0,5165	2,295	0,6243	2,340
4,3	0,3151	2,160	0,3907	2,186	0,4747	2,217	0,5708	2,256
4,4	0,2937	2,100	0,3632	2,122	0,4397	2,150	0,5265	2,184
4,5	0,2752	2,046	0,3396	2,066	0,4100	2,090	0,4891	2,121
4,6	0,2592	1,998	0,3192	2,016	0,3844	2,038	0,4564	2,065
4,7	0,2451	1,955	0,3014	1,971	0,3622	1,991	0,4288	2,015
4,8	0,2327	1,916	0,2857	1,930	0,3426	1,948	0,4048	1,970
4,9	0,2216	1,880	0,2717	1,893	0,3354	1,910	0,3836	1,930
5,0	0,2117	1,847	0,2592	1,860	0,3099	1,875	0,3648	1,893
5,1	0,2027	1,817	0,2480	1,829	0,2961	1,843	0,3479	1,859
5,2	0,1946	1,790	0,2378	1,800	0,2836	1,813	0,3328	1,829
5,3	0,1872	1,764	0,2285	1,774	0,2723	1,786	0,3191	1,800
5,4	0,1804	1,740	0,2201	1,749	0,2620	1,760	0,3066	1,774
5,5	0,1742	1,718	0,2183	1,726	0,2525	1,737	0,2952	1,749
5,6	0,1684	1,697	0,2052	1,705	0,2438	1,715	0,2848	1,773
5,7	0,1631	1,677	0,1986	1,685	0,2358	1,694	0,2752	1,705
5,8	0,1582	1,658	0,1925	1,666	0,2284	1,674	0,2663	1,685
5,9	0,1536	1,641	0,1868	1,648	0,2215	1,656	0,2581	1,666
6,1	0,1493	1,625	0,1815	1,631	0,2151	1,639	0,2504	1,648
6,2	0,1415	1,594	0,1719	1,600	0,2035	1,607	0,2366	1,615
6,3	0,1380	1,580	0,1676	1,586	0,1983	1,593	0,2304	1,600
6,4	0,1347	1,567	0,1635	1,572	0,1933	1,579	0,2245	1,586
6,5	0,1315	1,554	0,1596	1,559	0,1887	1,565	0,2190	1,573
6,6	0,1286	1,542	0,1560	1,547	0,1843	1,553	0,2138	1,560
6,7	0,1258	1,530	0,1525	1,535	0,1802	1,541	0,2089	1,547
6,8	0,1231	1,519	0,1492	1,524	0,1762	1,529	0,2043	1,535
6,9	0,1206	1,509	0,1461	1,513	0,1726	1,518	0,1999	1,524
7,0	0,1182	1,498	0,1432	1,502	0,1690	1,507	0,1957	1,513
7,1	0,1159	1,489	0,1404	1,492	0,1656	1,497	0,1918	1,503
7,2	0,1137	1,479	0,1377	1,483	0,1624	1,487	0,1880	1,493
7,3	0,1116	1,470	0,1352	1,474	0,1594	1,478	0,1844	1,483
7,4	0,1096	1,461	0,1327	1,465	0,1565	1,465	0,1810	1,474
7,5	0,1077	1,453	0,1304	1,456	0,1537	1,456	0,1777	1,465
7,6	0,1059	1,445	0,1282	1,448	0,1510	1,448	0,1746	1,457
7,7	0,1042	1,437	0,1260	1,440	0,1485	1,440	0,1716	1,448
7,8	0,1025	1,429	0,1240	1,432	0,1460	1,432	0,1687	1,440
7,9	0,1009	1,422	0,1220	1,425	0,1437	1,425	0,1660	1,433
8,0	0,0993	1,415	0,1201	1,418	0,1414	1,418	0,1633	1,425

Продолжение таблицы 116

α_4	α_3							
	0,45		0,50		0,55		0,60	
	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η
3,6	2,1390	3,705	3,1570	4,087				
3,7	1,6140	3,295	2,1880	3,540				
3,8	1,3020	3,011	1,6870	3,184	2,2840	3,424	3,3830	3,776
3,9	1,0950	2,801	1,3780	2,931	1,7830	3,105	2,4260	3,346
4,0	0,9470	2,637	1,1690	2,739	1,4690	2,872	1,9060	3,049
4,1	0,8363	2,505	1,0180	2,588	1,2530	2,694	1,5770	2,830
4,2	0,7503	2,396	0,9031	2,465	1,0960	2,552	1,3490	2,662
4,3	0,6814	2,304	0,8132	2,363	0,9748	2,436	1,1820	2,526
4,4	0,6251	2,226	0,7407	2,276	0,8802	2,338	1,0540	2,414
4,5	0,5780	2,157	0,6811	2,202	0,8033	2,255	0,9527	2,320
4,6	0,5382	2,097	0,6311	2,140	0,7399	2,136	0,9183	2,240
4,7	0,5040	2,044	0,5886	2,079	0,6865	2,120	0,8024	2,170
4,8	0,4744	1,997	0,5520	2,028	0,6410	2,065	0,7451	2,109
4,9	0,4484	1,954	0,5202	1,982	0,6017	2,015	0,6962	2,055
5,0	0,4254	1,915	0,4922	1,941	0,5675	1,971	0,6539	2,007
5,1	0,4050	1,880	0,4674	1,903	0,5274	1,931	0,6170	1,963
5,2	0,3866	1,847	0,4453	1,869	0,5106	1,894	0,5845	1,924
5,3	0,3696	1,817	0,4255	1,837	0,4868	1,860	0,5556	1,888
5,4	0,3547	1,789	0,4076	1,808	0,4653	1,830	0,5298	1,855
5,5	0,3411	1,764	0,3913	1,781	0,4459	1,801	0,5066	1,824
5,6	0,3286	1,740	0,3765	1,756	0,4283	1,775	0,4856	1,796
5,7	0,3172	1,718	0,3629	1,733	0,4122	1,750	0,4665	1,770
5,8	0,3066	1,697	0,3504	1,711	0,3975	1,728	0,4491	1,746
5,9	0,2967	1,677	0,3385	1,691	0,3840	1,706	0,4331	1,691
6,0	0,2879	1,659	0,3278	1,672	0,3714	1,686	0,4184	1,673
6,1	0,2794	1,642	0,3180	1,653	0,3598	1,6667	0,4049	1,653
6,2	0,2716	1,625	0,3088	1,636	0,3491	1,649	0,3923	1,664
6,3	0,2643	1,610	0,3002	1,620	0,3390	1,633	0,3806	1,647
6,4	0,2574	1,595	0,2921	1,605	0,3297	1,617	0,3697	1,630
6,5	0,2509	1,581	0,2846	1,591	0,3209	1,602	0,3595	1,614
6,6	0,2448	1,568	0,2775	1,577	0,3123	1,587	0,3500	1,599
6,7	0,2391	1,555	0,2709	1,564	0,3046	1,574	0,3410	1,585
6,8	0,2337	1,543	0,2646	1,551	0,2973	1,561	0,3326	1,572
6,9	0,2285	1,531	0,2586	1,539	0,2904	1,548	0,3247	1,559
7,0	0,2237	1,520	0,2530	1,528	0,2839	1,537	0,3172	1,547
7,1	0,2190	1,509	0,2476	1,517	0,2778	1,525	0,3101	1,453
7,2	0,2147	1,499	0,2426	1,506	0,2719	1,514	0,3034	1,524
7,3	0,2105	1,489	0,2377	1,496	0,2664	1,504	0,2967	1,513
7,4	0,2065	1,480	0,2331	1,487	0,2611	1,494	0,2907	1,503
7,5	0,2020	1,471	0,2287	1,477	0,2561	1,487	0,2849	1,493
7,6	0,1991	1,462	0,2246	1,468	0,2513	1,475	0,2795	1,483
7,7	0,1956	1,454	0,2206	1,460	0,2467	1,466	0,2742	1,447
7,8	0,1922	1,445	0,2167	1,451	0,2423	1,458	0,2692	1,465
7,9	0,1891	1,438	0,2131	1,443	0,2381	1,450	0,2645	1,457
8,0	0,1860	1,430	0,2095	1,436	0,2341	1,442	0,2599	1,448
8,2	0,1802	1,416	0,2029	1,421	0,2266	1,426	0,2514	1,433
8,4	0,1749	1,402	0,1968	1,407	0,2196	1,412	0,2435	1,418
8,6	0,1699	1,389	0,1912	1,394	0,2132	1,399	0,2362	1,405
8,8	0,1653	1,377	0,1859	1,381	0,2072	1,386	0,2294	1,392
9,0	0,1610	1,366	0,1810	1,370	0,2016	1,374	0,2231	1,380
9,2	0,1570	1,355	0,1764	1,359	0,1914	1,363	0,2172	1,368
9,4	0,1532	1,345	0,1721	1,348	0,1905	1,353	0,2117	1,357

Продолжение таблицы 116

α_4	α_3							
	0,65		0,70		0,75		0,80	
	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η
4,0	2,6210	3,294	4,1040	3,659				
4,1	2,0600	3,013	2,8860	3,269				
4,2	1,7050	2,804	2,2540	2,996				
4,3	1,4600	2,641	1,8600	2,791				
4,4	1,2800	2,510	1,5890	2,631				
4,5	1,1420	2,401	1,3920	2,502				
4,6	1,0330	2,309	1,2400	2,395	1,5220	2,503	1,9310	2,641
4,7	0,9434	2,231	1,1210	2,304	1,3530	2,395	1,6760	2,511
4,8	0,8699	2,162	1,0240	2,226	1,2210	2,305	1,4850	2,403
4,9	0,8083	2,100	0,9436	2,159	1,1140	2,227	1,3360	2,312
5,0	0,7550	2,049	0,8761	2,099	1,0250	2,160	1,2150	2,234
5,1	0,7092	2,001	0,8154	2,045	0,9509	2,100	1,1170	2,165
5,2	0,6693	1,958	0,7687	1,999	0,8877	2,048	1,0340	2,105
5,3	0,6341	1,919	0,7252	1,957	0,8331	2,000	0,9642	2,052
5,4	0,6028	1,884	0,6869	1,918	0,7586	1,958	0,9039	2,005
5,5	0,5749	1,851	0,6530	1,883	0,7419	1,918	0,8515	1,962
5,6	0,5498	1,821	0,6226	1,850	0,7067	1,884	0,8055	1,923
5,7	0,5270	1,794	0,5953	1,820	0,6735	1,851	0,7647	1,887
5,8	0,5063	1,768	0,5706	1,793	0,6337	1,821	0,7284	1,855
5,9	0,4875	1,744	0,5482	1,767	0,6168	1,794	0,6942	1,824
6,0	0,4701	1,722	0,5276	1,743	0,5924	1,777	0,6663	1,797
6,1	0,4542	1,701	0,5088	1,721	0,5700	1,744	0,6396	1,771
6,2	0,4395	1,681	0,4915	1,700	0,5496	1,722	0,6152	1,747
6,3	0,4258	1,663	0,4755	1,681	0,5308	1,701	0,5929	1,725
6,4	0,4131	1,645	0,4607	1,662	0,5134	1,682	0,5724	1,704
6,5	0,4013	1,629	0,4470	1,645	0,4973	1,663	0,5535	1,684
6,6	0,3903	1,613	0,4341	1,628	0,4824	1,646	0,5359	1,666
6,7	0,3799	1,598	0,4221	1,613	0,4684	1,629	0,5197	1,648
6,8	0,3702	1,584	0,4109	1,598	0,4554	1,614	0,5045	1,632
6,9	0,3611	1,571	0,4004	1,584	0,4433	1,599	0,4904	1,616
7,0	0,3524	1,558	0,3905	1,571	0,4318	1,585	0,4772	1,601
7,1	0,3443	1,546	0,3812	1,558	0,4211	1,572	0,4648	1,587
7,2	0,3366	1,.534	0,3723	1,546	0,4110	1,559	0,4531	1,573
7,3	0,3293	1,523	0,3640	1,534	0,4014	1,547	0,4421	1,561
7,4	0,3224	1,512	0,3561	1,523	0,3924	1,535	0,4318	1,548
7,5	0,3159	1,502	0,3486	1,512	0,3838	1,524	0,4220	1,537
7,6	0,3096	1,492	0,3415	1,502	0,3757	1,523	0,4127	1,525
7,7	0,3037	1,483	0,3347	1,492	0,3680	1,503	0,4039	1,515
7,8	0,2980	1,473	0,3283	1,483	0,3607	1,493	0,3956	1,504
7,9	0,2926	1,469	0,3221	1,474	0,3537	1,483	0,3876	1,494
8,0	0,2871	1,456	0,3163	1,465	0,3470	1,474	0,3800	1,485
8,1	0,2822	1,448	0,3106	1,456	0,3407	1,466	0,3729	1,476
8,2	0,2724	1,440	0,3053	1,448	0,3346	1,457	0,3660	1,467
8,3	0,2729	1,432	0,3001	1,440	0,3288	1,448	0,3594	1,458
8,4	0,2685	1,425	0,2952	1,433	0,3232	1,441	0,3531	1,450
8,5	0,2643	1,418	0,2904	1,425	0,3179	1,433	0,3471	1,442
8,6	0,2603	1,411	0,2859	1,418	0,3127	1,426	0,3413	1,435
8,7	0,2564	1,404	0,2812	1,411	0,3078	1,419	0,3358	1,427
8,8	0,2526	1,398	0,2770	1,404	0,3031	1,412	0,3305	1,420
8,9	0,2490	1,391	0,2730	1,398	0,2985	1,405	0,3253	1,413
9,0	0,2455	1,385	0,2691	1,392	0,2941	1,399	0,3204	1,406
9,2	0,2389	1,373	0,2617	1,379	0,2858	1,386	0,3111	1,393
9,4	0,2328	1,362	0,2548	1,368	0,2778	1,374	0,3025	1,381
9,6			0,2483	1,357	0,2706	1,363	0,2944	1,370
9,8			0,2422	1,347	0,2639	1,353	0,2868	1,359
10,0			0,2365	1,337	0,2575	1,343	0,2798	1,349

Продолжение таблицы 116

α_4	α_3							
	0,85		0,90		0,95		1,00	
	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η
4,6	2,6020	2,828	4,0290	3,093				
4,7	2,1670	2,662	3,0380	2,868				
4,8	1,8640	2,529	2,4730	2,694				
4,9	1,6410	2,418	2,1000	2,555				
5,0	1,4690	2,325	1,8320	2,441	2,4070	2,592	3,5380	2,799
5,1	1,3330	2,245	1,6290	2,344	2,0710	2,472	2,8320	2,641
5,2	1,2210	2,176	1,4700	2,262	1,8250	2,371	2,3860	2,512
5,3	1,1280	2,115	1,3420	2,191	1,6350	2,285	2,0720	2,406
5,4	1,0500	2,061	1,2360	2,128	1,4840	2,211	1,8390	2,315
5,5	0,9826	2,012	1,1470	2,073	1,3610	2,146	1,6570	2,237
5,6	0,9243	1,969	1,0710	2,023	1,2590	2,089	1,5100	2,170
5,7	0,8732	1,929	1,0060	1,979	1,1720	2,038	1,3900	2,110
5,8	0,8282	1,893	0,9485	1,939	1,0980	1,992	1,2890	2,057
5,9	0,7881	1,860	0,8982	1,902	1,0330	1,951	1,2030	2,009
6,0	0,7521	1,830	0,8536	1,868	1,9765	1,913	1,1290	1,967
6,1	0,7197	1,802	0,8138	1,837	0,9265	1,879	1,0650	1,928
6,2	0,6904	1,776	0,7780	1,809	0,8620	1,847	1,0080	1,892
6,3	0,6637	1,752	0,7456	1,782	0,8420	1,818	0,9581	1,860
6,4	0,6392	1,729	0,7161	1,758	0,8060	1,791	0,9132	1,830
6,5	0,6150	1,707	0,6892	1,735	0,7733	1,766	0,8729	1,802
6,6	0,5962	1,688	0,6646	1,713	0,7436	1,742	0,8364	1,776
6,7	0,5770	1,668	0,6419	1,693	0,7164	1,721	0,8033	1,752
6,8	0,5592	1,652	0,6209	1,678	0,6914	1,700	0,7730	1,730
6,9	0,5247	1,635	0,5961	1,656	0,6683	1,681	0,7453	1,708
7,0	0,5273	1,619	0,5835	1,639	0,6470	1,653	0,7198	1,689
7,1	0,5130	1,604	0,5666	1,623	0,6272	1,645	0,6962	1,670
7,2	0,4995	1,590	0,5509	1,608	0,6087	1,629	0,6744	1,653
7,3	0,4868	1,576	0,5362	1,594	0,5915	1,614	0,6541	1,636
7,4	0,4749	1,563	0,5224	1,580	0,5754	1,599	0,6352	1,620
7,5	0,4636	1,551	0,5094	1,567	0,5593	1,585	0,6171	1,605
7,6	0,4530	1,539	0,4972	1,555	0,5463	1,572	0,6010	1,591
7,7	0,4429	1,528	0,4857	1,543	0,5328	1,559	0,5855	1,578
7,8	0,4334	1,517	0,4747	1,531	0,5203	1,547	0,5709	1,565
7,9	0,4244	1,507	0,4644	1,520	0,5084	1,536	0,5571	1,552
8,0	0,4158	1,497	0,4546	1,510	0,4971	1,524	0,5442	1,541
8,1	0,4076	1,487	0,4452	1,500	0,4864	1,514	0,5309	1,529
8,2	0,3998	1,478	0,4364	1,490	0,4763	1,504	0,5203	1,519
8,3	0,3923	1,469	0,4279	1,481	0,4667	1,494	0,5092	1,508
8,4	0,3852	1,460	0,4199	1,472	0,4575	1,484	0,4987	1,498
8,5	0,3784	1,452	0,4122	1,463	0,4488	1,475	0,4888	1,489
8,6	0,3719	1,444	0,4048	1,455	0,4405	1,467	0,4793	1,479
8,7	0,3657	1,437	0,3978	1,447	0,4325	1,458	0,4702	1,471
8,8	0,3597	1,429	0,3910	1,439	0,4248	1,450	0,4616	1,462
8,9	0,3539	1,422	0,3845	1,431	0,4175	1,442	0,4533	1,454
9,0	0,3484	1,415	0,3783	1,424	0,4105	1,434	0,4454	1,446
9,2	0,3380	1,401	0,3666	1,410	0,3974	1,420	0,4305	1,431
9,4	0,3283	1,389	0,3558	1,397	0,3852	1,406	0,4169	1,416
9,6	0,3193	1,377	0,3457	1,385	0,3839	1,394	0,4042	1,403
9,8	0,3109	1,366	0,3363	1,373	0,3634	1,381	0,3925	1,390
10,0	0,3030	1,355	0,3275	1,362	0,3537	1,370	0,3816	1,379

Продолжение таблицы 116

α_4	α_3							
	1,05		1,10		1,15		1,20	
	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η
5,4	2,4030	2,450	2,5430	2,632				
5,5	2,0990	2,353	2,4780	2,505				
5,6	1,8710	2,270	2,4500	2,400				
5,7	1,6920	2,299	2,1490	2,311				
5,8	1,5470	2,136	1,9210	2,234	2,5350	2,362	3,8860	2,530
5,9	1,4280	2,080	1,7410	2,168	2,2240	2,278	3,1240	2,423
6,0	1,3270	2,031	1,5970	2,109	1,9900	2,206	2,6540	2,331
6,1	1,2410	1,986	1,4750	2,056	1,8050	2,143	2,3250	2,253
6,2	1,1670	1,945	1,3730	2,009	1,6560	2,087	2,0790	2,184
6,3	1,1020	1,908	1,2850	1,966	1,5310	2,037	1,8850	1,124
6,4	1,0440	1,875	1,2100	1,928	1,4270	1,992	1,7290	2,070
6,5	0,9933	1,843	1,1430	1,893	1,3370	1,951	1,5990	2,022
6,6	0,9477	1,815	1,0850	1,860	1,2590	1,914	1,4900	1,978
6,7	0,9066	1,788	1,0320	1,830	1,1900	1,880	1,3960	1,.939
6,8	0,8694	1,763	0,9857	1,803	1,1300	1,849	1,3150	1,903
6,9	0,8356	1,740	0,9435	1,777	1,0760	1,820	1,2430	1,870
7,0	0,8046	1,719	0,9053	1,753	1,0280	1,793	1,1800	1,840
7,1	0,7762	1,698	0,8705	1,731	1,9840	1,768	1,1240	1,811
7,2	0,7500	1,679	0,8386	1,710	0,9444	1,745	1,0740	1,785
7,3	0,7258	1,661	0,8093	1,690	0,9083	1,723	1,0280	1,761
7,4	0,7034	1,644	0,7823	1,671	0,8753	1,703	0,9871	1,738
7,5	0,6825	1,628	0,7573	1,654	0,8450	1,683	0,9495	1,717
7,6	0,6630	1,613	0,7342	1,637	0,8179	1,665	0,9151	1,697
7,7	0,6448	1,598	0,7126	1,622	0,7910	1,648	0,8834	1,678
7,8	0,6278	1,584	0,6924	1,607	0,7670	1,632	0,8542	1,660
7,9	0,6117	1,571	0,6736	1,593	0,7445	1,616	0,8272	1,644
8,0	0,5967	1,559	0,6559	1,579	0,7326	1,602	0,8020	1,628
8,1	0,5825	1,547	0,6393	1,566	0,7040	1,558	0,7786	1,613
8,2	0,5960	1,535	0,6237	1,554	0,6857	1,575	0,7568	1,598
8,3	0,5563	1,524	0,6089	1,542	0,6684	1,562	0,7464	1,585
8,4	0,5443	1,514	0,5950	1,531	0,6521	1,550	0,7172	1,571
8,5	0,5328	1,504	0,5818	1,520	0,6368	1,538	0,6991	1,559
8,6	0,5220	1,494	0,5693	1,510	0,6223	1,527	0,6822	1,547
8,7	0,5099	1,484	0,5574	1,500	0,6085	1,517	0,6661	1,536
8,8	0,5018	1,475	0,5461	1,490	0,5955	1,507	0,6510	1,525
8,9	0,4924	1,467	0,5354	1,481	0,5831	1,497	0,6366	1,514
9,0	0,4834	1,458	0,5251	1,472	0,5714	1,487	0,6230	1,504
9,1	0,4748	1,450	0,5154	1,463	0,5601	1,478	0,6101	1,495
9,2	0,4666	1,442	0,5060	1,455	0,5494	1,469	0,5978	1,485
9,3	0,4587	1,435	0,4971	1,447	0,5393	1,461	0,5861	1,476
9,4	0,4511	1,427	0,4885	1,440	0,5295	1,453	0,5749	1,468
9,5	0,4439	1,420	0,4803	1,432	0,5202	1,445	0,5642	1,459
9,6	0,4369	1,413	0,4724	1,425	0,5112	1,437	0,5540	1,451
9,7	0,4302	1,407	0,4648	1,418	0,5027	1,430	0,5443	1,443
9,8	0,4237	1,400	0,4576	1,411	0,4945	1,423	0,5349	1,436
9,9	0,4175	1,394	0,4506	1,404	0,4866	1,416	0,5260	1,428
10,0	0,4115	1,388	0,4438	1,398	0,4790	1,409	0,5174	1,421
10,2	0,4001	1,376	0,4311	1,386	0,4646	1,396	0,5012	1,408
10,4	0,3895	1,365	0,4192	1,374	0,4513	1,384	0,4862	1,395
10,6	0,3796	1,355	0,4082	1,363	0,4389	1,373	0,4723	1,383
10,8	0,3703	1,345	0,3978	1,353	0,4273	1,362	0,4593	1,372
11,0	0,3615	1,335	0,3881	1,343	0,4165	1,352	0,4472	1,361
11,2	0,3533	1,326	0,3789	1,334	0,4063	1,342	0,4358	1,351
11,4	0,3455	1,318	0,3703	1,325	0,3698	1,332	0,4252	1,341

Продолжение таблицы 116

α_4	α_3							
	1,25		1,30		1,35		1,40	
	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η
6,2	2,8230	2,309	5,1110	2,476				
6,3	2,4590	2,234	3,7120	2,378				
6,4	2,1920	2,168	3,0540	2,294				
6,5	1,9840	2,109	2,6350	2,221				
6,6	1,8170	2,057	2,3350	2,156				
6,7	1,6790	2,011	2,1060	2,100				
6,8	1,5630	1,969	1,9240	2,049	2,5250	2,151	3,9290	2,281
6,9	1,4640	1,930	1,7750	2,003	2,2590	2,094	3,2130	2,210
7,0	1,3790	1,895	1,6500	1,962	2,0550	2,044	2,7680	2,148
7,1	1,3030	1,863	1,5440	1,924	1,8900	1,999	2,4550	2,093
7,2	1,2370	1,833	1,4520	1,890	1,7520	1,958	2,2170	2,043
7,3	1,1780	1,806	1,3720	1,858	1,6360	1,921	2,0290	1,998
7,4	1,1250	1,780	1,3010	1,829	1,5370	1,887	1,8750	1,958
7,5	1,0770	1,756	1,2380	1,802	1,4500	1,856	1,7460	1,921
7,6	1,0340	1,734	1,1820	1,776	1,3740	1,827	1,6360	1,887
7,7	0,9945	1,713	1,1320	1,753	1,3080	1,800	1,5400	1,856
7,8	0,9584	1,693	1,0860	1,731	1,2460	1,775	1,4570	1,827
7,9	0,9251	1,675	1,0440	1,710	1,1920	1,751	1,3840	1,800
8,0	0,8945	1,657	1,0060	1,691	1,1430	1,730	1,3190	1,775
8,1	0,8661	1,640	0,9706	1,672	1,0990	1,709	1,2600	1,752
8,2	0,8397	1,625	0,9382	1,655	1,0580	1,690	1,2070	1,730
8,3	0,8152	1,610	0,9083	1,639	1,0200	1,671	1,1590	1,709
8,4	0,7923	1,596	0,8805	1,623	0,9860	1,654	1,1150	1,690
8,5	0,7709	1,582	0,8546	1,608	0,9542	1,638	1,0750	1,672
8,6	0,7507	1,569	0,8304	1,594	0,9246	1,623	1,0380	1,655
8,7	0,7318	1,557	0,8078	1,581	0,8972	1,608	1,0040	1,639
8,8	0,7140	1,545	0,7866	1,568	0,8716	1,594	0,9729	1,623
8,9	0,6972	1,534	0,7667	1,556	0,8477	1,580	0,9436	1,608
9,0	0,6813	1,523	0,7480	1,544	0,8252	1,568	0,9136	1,594
9,1	0,6663	1,513	0,7303	1,533	0,8042	1,566	0,8908	1,581
9,2	0,6520	1,503	0,7136	1,522	0,7843	1,440	0,8669	1,568
9,3	0,6385	1,493	0,6977	1,512	0,7656	1,533	0,8445	1,556
9,4	0,6256	1,484	0,6827	1,502	0,7480	1,522	0,8234	1,545
9,5	0,6133	1,475	0,6685	1,492	0,7312	1,512	0,8036	1,534
9,6	0,6016	1,466	0,6549	1,483	0,7154	1,502	0,7848	1,523
9,7	0,5904	1,458	0,6420	1,474	0,7003	1,492	0,7671	1,513
9,8	0,5797	1,450	0,6297	1,466	0,6860	1,474	0,7503	1,503
9,9	0,5695	1,442	0,6180	1,458	0,6724	1,466	0,7343	1,493
10,0	0,5597	1,435	0,6067	1,450	0,6594	1,456	0,7192	1,484
10,2	0,5413	1,420	0,5857	1,434	0,6352	1,450	0,6910	1,467
10,4	0,5243	1,407	0,5664	1,420	0,6131	1,435	0,6654	1,450
10,6	0,5086	1,394	0,5485	1,407	0,5927	1,420	0,6420	1,435
10,8	0,4940	1,382	0,5320	1,394	0,5739	1,407	0,6205	1,421
11,0	0,4804	1,371	0,5167	1,382	0,5566	1,394	0,6007	1,408
11,2	0,4678	1,360	0,5025	1,371	0,5404	1,383	0,5823	1,395
11,4	0,4559	1,350	0,4891	1,360	0,5254	1,371	0,5653	1,383
11,6							0,5495	1,372
11,8							0,5347	1,361
12,0							0,5209	1,351

Продолжение таблицы 116

α_4	α_3							
	1,45		1,50		1,55		1,60	
	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η
7,4	2,4300	2,046	3,6690	2,157				
7,5	2,2020	2,001	3,0880	2,102				
7,6	2,0240	1,960	2,7050	2,051				
7,7	1,8770	1,923	2,4260	2,006				
7,8	1,7520	1,889	2,2110	1,965				
7,9	1,646	1,858	2,0370	1,928				
8,0	1,5540	1,829	1,8930	1,894	2,4590	1,974	3,8130	2,074
8,1	1,4730	1,802	1,7720	1,862	2,2400	1,936	3,1920	2,028
8,2	1,4010	1,777	1,6670	1,833	2,0670	1,901	2,7950	1,985
8,3	1,3370	1,754	1,5760	1,806	1,9240	1,869	2,5100	1,947
8,4	1,2790	1,732	1,4960	1,781	1,8030	1,840	2,2910	1,911
8,5	1,2270	1,711	1,4250	1,758	1,6980	1,812	2,1150	1,879
8,6	1,1800	1,692	1,3610	1,736	1,6070	1,787	1,9690	1,849
8,7	1,1360	1,674	1,3040	1,715	1,5270	1,763	1,8460	1,821
8,8	1,0960	1,657	1,2520	1,695	1,4560	1,741	1,7400	1,795
8,9	1,0600	1,640	2,2040	1,677	1,3920	1,720	1,6470	1,771
9,0	1,0260	1,625	1,1610	1,660	1,3340	1,700	1,5660	1,748
9,1	0,9943	1,610	1,1210	1,643	1,2820	1,682	1,4930	1,727
9,2	0,9651	1,596	1,0840	1,628	1,2340	1,664	1,4280	1,707
9,3	0,9377	1,583	1,0500	1,613	1,1900	1,648	1,3700	1,688
9,4	0,9122	1,570	1,0190	1,599	1,1500	1,632	1,3160	1,670
9,5	0,8882	1,558	0,9892	1,586	1,1120	1,617	1,2680	1,653
9,6	0,8657	1,546	0,9616	1,573	1,0780	1,603	1,2230	1,637
9,7	0,8445	1,535	0,9359	1,560	1,0460	1,589	1,1820	1,622
9,8	0,8245	1,524	0,9117	1,549	1,0160	1,576	1,1440	1,608
9,9	0,8056	1,514	0,8889	1,538	0,9881	1,564	1,1090	1,594
10,0	0,7877	1,504	0,8675	1,527	0,9619	1,552	1,0760	1,581
10,2	0,7546	1,485	0,8280	1,506	0,9142	1,530	1,0170	1,557
10,4	0,7247	1,468	0,7927	1,488	0,8718	1,509	0,9655	1,534
10,6	0,6975	1,452	0,7608	1,470	0,8338	1,491	0,9196	1,513
10,8	0,6727	1,437	0,7318	1,454	0,7996	1,473	0,8785	1,494
11,0	0,6499	1,422	0,7053	1,439	0,7686	1,456	0,8416	1,476
11,2	0,6289	1,409	0,6811	1,424	0,7403	1,441	0,8083	1,460
11,4	0,6095	1,396	0,6588	1,411	0,7145	1,427	0,7780	1,444
11,6	0,5915	1,.384	0,6383	1,398	0,6907	1,413	0,7503	1,430
11,8	0,5748	1,373	0,6192	1,386	0,6688	1,400	0,7248	1,416
12,0	0,5592	1,365	0,6016	1,375	0,6486	1,388	0,7014	1,403
12,2					0,6297	1,377	0,6798	1,391
12,4					0,6122	1,366	0,6598	1,379
12,6					0,5959	1,356	0,6411	1,369
12,8					0,5806	1,346	0,6237	1,358
13,0					0,5662	1,337	0,6074	1,348

Напомним, что

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2; \quad \operatorname{ch} z = \frac{e^z + e^{-z}}{2}; \quad \operatorname{sh} z = \frac{e^z - e^{-z}}{2}.$$

Квантили u_α^x эмпирического семейства S_U выражаются через квантили u_α стандартного нормального распределения с помощью формулы

$$u_\alpha^x = \varepsilon + \lambda \operatorname{sh}\left(\frac{u_\alpha - \gamma}{\eta}\right).$$

Продолжение таблицы 116

α_4	α_3							
	1,65		1,70		1,75		1,80	
	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η
8,6	2,6010	1,925	4,6030	2,018				
8,7	2,3650	1,891	3,5820	1,978				
8,8	2,1810	1,860	3,0610	1,940				
8,9	2,0300	1,831	2,7190	1,906				
9,0	1,9020	1,805	2,4670	1,874				
9,1	1,7930	1,780	2,2690	1,844				
9,2	1,6970	1,757	2,1080	1,817				
9,3	1,6130	1,735	1,9730	1,792				
9,4	1,5390	1,715	1,8580	1,768	2,3820	1,831	3,6790	1,910
9,5	1,4720	1,696	1,7580	1,745	2,2080	1,805	3,1470	1,879
9,6	1,4120	1,678	1,6700	1,725	2,0620	1,781	2,7930	1,849
9,7	1,3570	1,660	1,5920	1,705	1,9380	1,758	2,5350	1,822
9,8	1,3070	1,644	1,5230	1,686	1,8310	1,736	2,3350	1,796
9,9	1,2620	1,629	1,4600	1,669	1,7380	1,716	2,1720	1,772
10,0	1,2190	1,614	1,4030	1,652	1,6560	1,697	2,0360	1,750
10,2	1,1440	1,587	1,3040	1,621	1,5160	1,662	1,8190	1,709
10,4	1,0790	1,562	1,2200	1,594	1,4020	1,631	1,6520	1,673
10,6	1,0220	1,539	1,1480	1,568	1,3070	1,602	1,5180	1,640
10,8	0,9720	1,518	1,0850	1,545	1,2260	1,576	1,4080	1,631
11,0	0,9274	1,490	1,0300	1,524	1,1560	1,552	1,3160	1,584
11,2	0,8874	1,481	0,9811	1,504	1,0950	1,530	1,2360	1,560
11,4	0,8513	1,464	0,9375	1,485	1,0410	1,510	1,1680	1,538
11,6	0,8186	1,448	0,8983	1,468	0,9928	1,491	1,1080	1,517
11,8	0,7888	1,433	0,8628	1,452	0,9499	1,474	1,0540	1,498
12,0	0,7615	1,419	0,8306	1,437	0,9112	1,458	1,0070	1,480
12,2	0,7364	1,406	0,8011	1,423	0,8761	1,442	0,9644	1,463
12,4	0,7132	1,394	0,7740	1,410	0,8441	1,428	0,9260	1,448
12,6	0,6918	1,382	0,7491	1,398	0,8148	1,415	0,8910	1,433
12,8	0,6719	1,371	0,7261	1,386	0,7879	1,402	0,8592	1,420
13,0	0,6533	1,361	0,7047	1,375	0,7630	1,390	0,8299	1,407
13,2					0,7400	1,379	0,8030	1,394
13,4					0,7187	1,368	0,7781	1,383
13,6					0,6988	1,358	0,7551	1,372
13,8					0,6802	1,348	0,7336	1,362
14,0					0,6628	1,339	0,7136	1,352
14,2					0,6464	1,330	0,6949	1,342
14,4					0,6311	1,321	0,6774	1,333
14,6					0,6166	1,313	0,6609	1,325
14,8					0,6029	1,305	0,6454	1,316
15,0					0,5900	1,298	0,6308	1,308

Задача 172. *В результате наблюдений над случайной величиной x получены следующие значения ($n = 50$):*

11	16	18	21	21	23	24	24	25	26
27	28	28	28	28	29	29	30	31	31
31	32	32	32	33	33	34	34	35	35
35	36	36	37	37	38	38	39	40	40
40	41	42	42	45	46	48	55	60	62

Найти аппроксимирующее распределение Джонсона, оценить его параметры и вычислить значение $x = x^*$, для которого $\alpha = \mathbf{P}(x \leqslant x^*) = 0{,}95$.

Окончание таблицы 116

α_4	α_3							
	1,85		1,90		1,95		2,00	
	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η	$-\gamma$	η
10,2	2,3100	1,766	3,4750	1,834				
10,4	2,0280	1,723	2,7260	1,784				
10,6	1,8200	1,686	2,3130	1,740				
10,8	1,6580	1,652	2,0360	1,701				
11,0	1,5280	1,622	1,8320	1,666	2,3310	1,717	3,6120	1,780
11,2	1,4200	1,594	1,6730	1,634	2,0580	1,681	2,8040	1,737
11,4	1,3290	1,569	1,5440	1,605	1,8550	1,648	2,3770	1,698
11,6	1,2510	1,546	1,4380	1,580	1,6960	1,618	2,0960	1,664
11,8	1,1830	1,525	1,3470	1,556	1,5680	1,591	1,8890	1,633
12,0	1,1240	1,505	1,2700	1,534	1,4610	1,567	1,7280	1,605
12,2	1,0710	1,487	1,2020	1,514	1,3700	1,544	1,5980	1,579
12,4	1,0240	1,470	1,1430	1,495	1,2930	1,523	1,4900	1,555
12,6	0,9811	1,454	1,0900	1,477	1,2250	1,504	1,3990	1,534
12,8	0,9427	1,439	1,0430	1,461	1,1650	1,486	1,3200	1,514
13,0	0,9078	1,425	1,0000	1,446	1,1120	1,469	1,2510	1,495
13,2	0,8758	1,412	0,9614	1,431	1,0640	1,453	1,1900	1,478
13,4	0,8465	1,400	0,9263	1,418	1,0210	1,438	1,1360	1,461
13,6	0,8195	1,388	0,8941	1,405	0,9822	1,425	1,0880	1,446
13,8	0,7945	1,377	0,8646	1,393	0,9466	1,412	1,0450	1,432
14,0	0,7712	1,366	0,8373	1,382	0,9141	1,399	1,0050	1,419
14,2	0,7496	1,356	0,8121	1,371	0,8842	1,388	0,9690	1,406
14,4	0,7295	1,346	0,7886	1,361	0,8566	1,377	0,9359	1,394
14,6	0,7106	1,337	0,7668	1,351	0,8310	1,366	0,9055	1,383
14,8	0,6929	1,328	0,7464	1,342	0,8072	1,356	0,8774	1,372
15,0	0,6763	1,320	0,7273	1,333	0,7851	1,346	0,8514	1,362

Находим

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = 33{,}72; \quad s^2 = 100{,}858; \quad s = 10{,}0428; \quad \sum_{i=1}^{n} (x_i - \bar{x})^3 = 28396{,}44481;$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^4 = 1947356{,}956; \quad \alpha_3 = 0{,}5607; \quad \alpha_4 = 3{,}8287.$$

Так как $\alpha_4 = 3{,}8287 > 3 \cdot \left(1 + 0{,}641 \cdot \alpha_3^2\right) = 3{,}604$, применима аппроксимация распределением из семейства S_U Джонсона. Для $\alpha_4 = 3{,}83$ и $\alpha_3 = 0{,}56$ из табл. 116 (используя аппроксимацию) находим оценки $\hat{\gamma} = -2{,}1$ и $\hat{\eta} = 3{,}30$.

Далее вычисляем

$$\omega = \exp\left(\frac{1}{\hat{\eta}^2}\right) = e^{0{,}091827} = 1{,}096175;$$

$$\operatorname{ch}\left(\frac{2\hat{\gamma}}{\hat{\eta}}\right) = \operatorname{ch}\left[\frac{2 \cdot (-2{,}11)}{3{,}30}\right] = \operatorname{ch}(-1{,}27878) = 1{,}9353145;$$

$$\operatorname{sh}\left(\frac{\hat{\gamma}}{\hat{\eta}}\right) = \operatorname{sh}\left(-\frac{2{,}11}{3{,}30}\right) = -0{,}68386.$$

Тогда

$$\hat{\lambda} = \frac{10{,}0428}{\sqrt{\frac{1}{2} \cdot (1{,}09618 - 1) \cdot (1{,}09618 \cdot 1{,}9353 + 1)}} = 25{,}9209;$$

$$\hat{\varepsilon} = 33{,}77 + 25{,}9209 \cdot \sqrt{1{,}09618} \cdot (-0{,}68386) = 15{,}2108.$$

Плотность аппроксимирующего распределения семейства S_U Джонсона имеет вид

$$f(x) = \frac{3{,}3}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{(x-15{,}21)^{+}\,25{,}921^2}} \times$$
$$\times \exp\left\{ -\frac{1}{2} \cdot \left(-2{,}11 + 3{,}3 \cdot \ln\left\{ \frac{x-15{,}21}{25{,}921} + \sqrt{\left(\frac{x-15{,}21}{25{,}921}\right)^2 + 1} \right\} \right)^2 \right\}.$$

Для того, чтобы найти x^*, для которого $\alpha = \mathbf{P}(x \leqslant x^*) = 0{,}95$, учитывая, что $u_\alpha = u_{0,95} = 1{,}645$, и полагая по определению $x^* = u_\alpha^x = u_{0,95}^x$, находим

$$u_{0,95}^x = 15{,}16 + 25{,}92 \cdot \operatorname{sh}\left(\frac{1{,}645 + 2{,}11}{3{,}30}\right) = 51{,}49.$$

Следовательно, $x \leqslant 51{,}49$ с вероятностью $\alpha = 0{,}95$.

3.6.2. Кривые распределений Пирсона

Наибольшее распространение для аппроксимации эмпирических распределений получили кривые Пирсона. Плотность вероятности $f(x)$, график которой принадлежит семейству Пирсона, является решением дифференциального уравнения [10]

$$\frac{dy}{y} = \frac{x+b}{c_0 + c_1 x + c_2 x^2}.$$

Постоянные b, c_0, c_1, c_2 выражаются через первые четыре момента распределения (математическое ожидание, дисперсия, коэффициенты асимметрии α_3 и эксцесса α_4):

$$c_0 = -\sigma^2 \frac{K+1}{K-2}; \quad c_1 = -b = \frac{\alpha_3 \sigma}{2} \frac{K+2}{K-1}; \quad c_2 = \frac{1}{2K},$$

где $K = \dfrac{6\left(\alpha_4 - \alpha_3^2 - 1\right)}{3\alpha_3^2 - 2\alpha_4 + 6}$.

Выбор из семейства кривых Пирсона такой кривой, у которой первые четыре момента совпадают с выборочными моментами, определенными по экспериментальным данным, составляет содержание задачи подбора эмпирической кривой для распределения вероятностей случайной величины.

Тип кривой из семейства Пирсона определяется значением показателя χ [10]

$$\chi = \frac{\alpha_3^2(\alpha_4+3)^2}{4(4\alpha_4 - 3\alpha_3^2)(2\alpha_4 - 3\alpha_3^2 - 6)} = -\frac{\alpha_3^2}{16} \frac{(K+2)^2}{(K+1)}.$$

На практике различают 7 основных типов кривых Пирсона, которым соответствуют различные значения χ, а следовательно, α_3 и α_4. Графики для определения типа кривой Пирсона по значениям α_3 и α_4 приведены в [25], по показателю χ — в [10].

Так как достаточно эффективные и стабильные оценки моментов α_3 и α_4 распределения достигаются при значительных объемах выборок, то анализу подвергаются, как правило, статистические ряды, разбитые на интервалы равной длины $c = x_{i+1} - x_i$, т. е. статистический ряд представляется своей эмпирической гистограммой.

В этом случае границы интервалов разбиения (x_i, x_{i+1}) фиксируются, а случайной величиной является количество выборочных значений данных n_i, попавших в тот или иной интервал.

Далее для всех кривых Пирсона переменная представлена в виде $x = \dfrac{x^* - \hat{x}}{c}$, где x^* — реальное значение переменной; $\hat{x}$ — мода распределения (значение случайной величины, соответствующее максимуму плотности распределения); c — длина интервала, на которые разбит эмпирический статистический ряд.

Напомним порядок вычисления моментов распределения, заданного эмпирической гистограммой. Гистограмма представлена совокупностью пар

$$\begin{pmatrix} x_1 \\ n_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ n_2 \end{pmatrix}, \begin{pmatrix} x_3 \\ n_3 \end{pmatrix}, \ldots, \begin{pmatrix} x_i \\ n_i \end{pmatrix}, \ldots, \begin{pmatrix} x_N \\ n_N \end{pmatrix},$$

где x_i — середина i-го интервала разбиения; n_i — количество данных, попавших в i-й интервал; N — количество интервалов разбиения гистограммы.

Первые четыре момента распределения подсчитываются по формулам $\left(n = \sum\limits_{i=1}^{N} n_i\right)$: $m_k = \dfrac{1}{n} \sum\limits_{i=1}^{N} x_i^k n_i$, $k = 1, 2, 3, 4$.

Часто вместо переменной x удобнее использовать переменную $\tilde{x} = \dfrac{x - x_0}{c}$, где x_0 — условное начало отсчета (как правило, это величина, соответствующая интервалу разбиения с наибольшей частотой n_i). Тогда $\bar{x} = x_0 + m_1 c$.

Очевидно, что $cm_1 = \bar{x}$, т. е. первый начальный момент равен среднему значению. Дисперсия σ^2 выражается через начальные моменты формулой $\sigma^2 = m_2 - m_1^2$, или $\sigma = \sqrt{m_2 - m_1^2}$. Переход к метрической величине осуществляется по формуле $\sigma' = c\sigma$.

Коэффициенты асимметрии и эксцесса выражаются через начальные моменты следующим образом:

$$\alpha_3 = \frac{m_3 - 3m_2 m_1 + 2m_1^3}{(m_2 - m_1^2)^{\frac{3}{2}}}; \quad \alpha_4 = \frac{m_4 - 4m_3 m_1 + 6m_2 m_1^3 - 3m_1^4}{(m_2 - m_1^2)^2}.$$

Легко видеть, что для негруппированного ряда $c = 1$, $n_i = 1$. Порядок вычисления m_k будет продемонстрирован ниже.

Аппроксимация распределением из семейства Пирсона позволяет подыскать подходящую кривую для описания плотности распределения эмпирических данных, что необходимо для выявления основных характеристик распределения (его ход, симметричность, поведение на хвостах). Однако найти квантиль этого распределения по аналитической формуле плотности распределения — задача непростая. Эту задачу можно решить без подбора распределения, достаточно найти оценки α_3 и α_4 и воспользоваться известными таблицами квантилей распределения Пирсона (табл. 117) для нормированной переменной $y = \dfrac{x - \bar{x}}{s}$ (тогда истинная квантиль равна $x_\beta = \bar{x} + s y_\alpha$).

3.6.2.1. Кривые Пирсона типа I

Для кривых этого семейства $\chi < 0$. Уравнение кривой имеет вид

$$y = \lambda \left(1 + \frac{\tilde{\tilde{x}}}{l_1}\right)^{q_1} \left(1 - \frac{\tilde{\tilde{x}}}{l_2}\right)^{q_2}, \quad -l_1 < x < l_2,$$

где

$$q_1(q_2) = \left\{ \frac{1}{2} \left[(K - 2) \mp K(K + 2) \frac{\alpha_3}{4\sqrt{(K+1)(1-\chi)}} \right] \right\};$$

Таблица 117

Значения квантилей $y_\alpha = (x_\alpha - \bar{x})/s$ нормированного распределения Пирсона

(при $\alpha_3 > 0$ следует поменять знак у нижних квантилей на обратный)

α_4	α_3^2														
	0,00	0,01	0,03	0,05	0,10	0,15	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00
$\alpha = 0,05$															
1,8	1,56														
2,0	1,61	1,56	1,51	1,47											
2,2	1,64	1,59	1,55	1,46	1,40	1,35									
2,4	1,65	1,61	1,58	1,55	1,50	1,45	1,41	1,33							
2,6	1,65	1,61	1,59	1,57	1,53	1,49	1,45	1,38	1,30						
2,8	1,65	1,62	1,59	1,57	1,54	1,51	1,48	1,42	1,35	1,29					
3,0	1,64	1,62	1,59	1,58	1,55	1,52	1,49	1,44	1,39	1,33	1,27				
3,2	1,64	1,61	1,59	1,58	1,55	1,53	1,50	1,46	1,42	1,37	1,31	1,25	1,19		
3,4	1,64	1,61	1,59	1,58	1,55	1,53	1,51	1,47	1,43	1,39	1,35	1,30	1,24	1,19	
3,6	1,63	1,61	1,59	1,58	1,55	1,53	1,52	1,48	1,45	1,41	1,37	1,33	1,28	1,23	1,18
3,8	1,63	1,60	1,59	1,58	1,55	1,54	1,52	1,49	1,46	1,42	1,39	1,35	1,31	1,27	1,23
4,0	1,62	1,60	1,59	1,57	1,55	1,54	1,52	1,49	1,46	1,43	1,40	1,37	1,34	1,30	1,26
4,2	1,62	1,60	1,58	1,57	1,55	1,54	1,52	1,49	1,47	1,44	1,42	1,39	1,36	1,32	1,28
4,4	1,62	1,60	1,58	1,57	1,55	1,54	1,52	1,50	1,47	1,45	1,42	1,40	1,37	1,34	1,31
4,6	1,61	1,59	1,58	1,57	1,55	1,54	1,52	1,50	1,48	1,45	1,43	1,41	1,38	1,35	1,33
4,8	1,61	1,59	1,58	1,57	1,55	1,53	1,52	1,50	1,48	1,46	1,44	1,41	1,39	1,37	1,34
5,0	1,60	1,58	1,57	1,56	1,55	1,53	1,52	1,50	1,48	1,46	1,44	1,42	1,40	1,38	1,35
$\alpha = 0,95$															
1,8	1,56														
2,0	1,61	1,66	1,70	1,72											
2,2	1,64	1,68	1,71	1,74	1,77	1,80	1,83								
2,4	1,65	1,69	1,71	1,74	1,77	1,80	1,83	1,87							
2,6	1,65	1,68	1,71	1,73	1,76	1,79	1,81	1,86	1,90						
2,8	1,65	1,68	1,70	1,72	1,75	1,77	1,80	1,84	1,88	1,92					
3,0	1,64	1,67	1,69	1,71	1,74	1,76	1,78	1,82	1,86	1,90	1,93				
3,2	1,64	1,67	1,69	1,70	1,73	1,75	1,77	1,80	1,84	1,87	1,91	1,94	1,98		
3,4	1,64	1,66	1,68	1,69	1,72	1,74	1,76	1,79	1,82	1,85	1,88	1,92	1,95	1,98	
3,6	1,63	1,65	1,67	1,68	1,71	1,63	1,74	1,77	1,80	1,83	1,86	1,89	1,92	1,95	1,98
3,8	1,63	1,65	1,66	1,68	1,70	1,72	1,73	1,76	1,79	1,82	1,84	1,87	1,90	1,93	1,96
4,0	1,62	1,64	1,66	1,67	1,69	1,71	1,72	1,75	1,78	1,80	1,83	1,85	1,88	1,90	1,93
4,2	1,62	1,64	1,65	1,66	1,68	1,70	1,71	1,74	1,76	1,79	1,81	1,83	1,86	1,88	1,91
4,4	1,62	1,63	1,65	1,66	1,68	1,69	1,70	1,73	1,75	1,78	1,80	1,82	1,84	1,86	1,89
4,6	1,61	1,63	1,64	1,65	1,67	1,68	1,70	1,72	1,74	1,76	1,78	1,80	1,82	1,84	1,85
4,8	1,61	1,62	1,64	1,65	1,66	1,68	1,69	1,71	1,73	1,75	1,77	1,79	1,81	1,83	1,85
5,0	1,60	1,62	1,63	1,64	1,66	1,67	1,68	1,71	1,73	1,74	1,76	1,78	1,80	1,82	1,84

Продолжение таблицы 117

α_4	α_3^2														
	0,00	0,01	0,03	0,05	0,10	0,15	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00
$\alpha = 0{,}025$															
1,8	1,65														
2,0	1,76	1,68	1,62	1,56											
2,2	1,83	1,76	1,71	1,66	1,57	1,49	1,41								
2,4	1,88	1,82	1,77	1,73	1,65	1,58	1,51	1,39							
2,6	1,92	1,86	1,82	1,78	1,71	1,64	1,58	1,47	1,37						
2,8	1,94	1,89	1,85	1,82	1,76	1,70	1,65	1,55	1,45	1,35					
3,0	1,96	1,91	1,87	1,84	1,79	1,74	1,69	1,60	1,52	1,42	1,33				
3,2	1,97	1,93	1,89	1,86	1,81	1,77	1,72	1,65	1,57	1,49	1,40	1,32	1,24		
3,4	1,98	1,94	1,90	1,88	1,83	1,79	1,75	1,68	1,61	1,54	1,46	1,39	1,31	1,23	
3,6	1,99	1,95	1,91	1,89	1,85	1,81	1,77	1,71	1,65	1,58	1,51	1,44	1,38	1,30	1,23
3,8	1,99	1,95	1,92	1,90	1,86	1,82	1,79	1,73	1,67	1,62	1,56	1,49	1,43	1,36	1,29
4,0	1,99	1,96	1,93	1,91	1,87	1,84	1,81	1,75	1,70	1,64	1,59	1,53	1,47	1,41	1,35
4,2	2,00	1,96	1,93	1,91	1,88	1,84	1,82	1,76	1,72	1,67	1,62	1,56	1,51	1,45	1,40
4,4	2,00	1,96	1,94	1,92	1,88	1,85	1,83	1,78	1,73	1,69	1,64	1,59	1,54	1,49	1,44
4,6	2,00	1,96	1,94	1,92	1,89	1,86	1,83	1,79	1,75	1,70	1,66	1,62	1,57	1,52	1,47
4,8	2,00	1,97	1,94	1,93	1,89	1,87	1,84	1,80	1,76	1,72	1,68	1,64	1,59	1,55	1,50
5,0	2,00	1,97	1,94	1,93	1,90	1,87	1,85	1,81	1,77	1,73	1,69	1,65	1,61	1,57	1,53
$\alpha = 0{,}975$															
1,8	1,65														
2,0	1,76	1,82	1,86	1,89											
2,2	1,83	1,89	1,93	1,96	2,00	2,04	2,06								
2,4	1,88	1,94	1,98	2,01	2,05	2,08	2,11	2,25							
2,6	1,92	1,97	2,01	2,03	2,08	2,11	2,14	2,18	2,22						
2,8	1,94	1,99	2,03	2,05	2,09	2,13	2,15	2,20	2,24	2,27					
3,0	1,96	2,01	2,04	2,06	2,10	2,13	2,16	2,21	2,25	2,28	2,32				
3,2	1,97	2,02	2,05	2,07	2,11	2,14	2,16	2,21	2,25	2,29	2,32	2,35	2,38		
3,4	1,98	2,02	2,05	2,07	2,11	2,14	2,16	2,21	2,25	2,28	2,32	2,35	2,38	2,41	
3,6	1,99	2,02	2,05	2,07	2,11	2,14	2,16	2,20	2,24	2,28	2,31	2,34	2,37	2,41	2,44
3,8	1,99	2,03	2,05	2,07	2,11	2,13	2,16	2,20	2,24	2,27	2,30	2,33	2,36	2,40	2,43
4,0	2,00	2,03	2,05	2,07	2,11	2,13	2,15	2,19	2,23	2,26	2,29	2,32	2,35	2,38	2,41
4,2	2,00	2,03	2,05	2,07	2,10	2,13	2,15	2,19	2,22	2,25	2,28	2,31	2,34	2,37	2,40
4,4	2,00	2,03	2,05	2,07	2,10	2,13	2,15	2,18	2,22	2,25	2,28	2,31	2,33	2,36	2,39
4,6	2,00	2,03	2,05	2,07	2,10	2,12	2,14	2,18	2,21	2,24	2,27	2,30	2,32	2,35	2,38
4,8	2,00	2,03	2,05	2,07	2,10	2,12	2,14	2,17	2,21	2,23	2,26	2,29	2,31	2,33	2,36
5,0	2,00	2,03	2,05	2,07	2,09	2,12	2,14	2,17	2,20	2,23	2,25	2,28	2,30	2,33	2,35

Окончание таблицы 117

α_4	α_3^2: 0,00	0,01	0,03	0,05	0,10	0,15	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00
	$\alpha = 0{,}01$														
1,8	1,70														
2,0	1,87	1,77	1,69	1,62											
2,2	2,01	1,91	1,83	1,76	1,64	1,55	1,45								
2,4	2,12	2,03	1,95	1,89	1,77	1,68	1,59	1,43							
2,6	2,21	2,12	2,05	1,99	1,88	1,79	1,70	1,55	1,41						
2,8	2,27	2,19	2,13	2,08	1,98	1,89	1,81	1,66	1,52	1,39					
3,0	2,33	2,25	2,19	2,14	2,05	1,97	1,90	1,76	1,62	1,50	1,38				
3,2	2,37	2,29	2,24	2,19	2,11	2,03	1,96	1,84	1,71	1,59	1,48	1,37	1,26		
3,4	2,40	2,33	2,28	2,24	2,16	2,09	2,02	1,90	1,79	1,68	1,57	1,46	1,36	1,26	
3,6	2,43	2,36	2,31	2,27	2,20	2,13	2,07	1,96	1,86	1,76	1,65	1,55	1,45	1,35	1,26
3,8	2,45	2,39	2,34	2,30	2,23	2,17	2,11	2,01	1,91	1,82	1,72	1,62	1,53	1,43	1,34
4,0	2,47	2,41	2,36	2,33	2,26	2,20	2,15	2,05	1,96	1,87	1,78	1,69	1,60	1,51	1,42
4,2	2,49	2,43	2,38	2,35	2,28	2,23	2,18	2,09	2,00	1,92	1,83	1,75	1,66	1,58	1,49
4,4	2,50	2,44	2,40	2,37	2,31	2,25	2,21	2,12	2,04	1,96	1,88	1,80	1,72	1,64	1,56
4,6	2,51	2,46	2,42	2,38	2,32	2,27	2,23	2,15	2,07	2,00	1,92	1,80	1,77	1,70	1,62
4,8	2,52	2,47	2,43	2,40	2,34	2,29	2,25	2,17	2,10	2,03	1,96	1,88	1,81	1,74	1,67
5,0	2,53	2,48	2,44	2,41	2,36	2,31	2,27	2,19	2,12	2,06	1,99	1,92	1,85	1,79	1,72
	$\alpha = 0{,}99$														
1,8	1,70														
2,0	1,87	1,95	2,00	2,03											
2,2	2,01	2,10	2,15	2,18	2,22	2,24	2,25								
2,4	2,12	2,20	2,25	2,28	2,33	2,36	2,38	2,40							
2,6	2,21	2,28	2,33	2,36	2,42	2,45	2,48	2,51	2,52						
2,8	2,27	2,34	2,39	2,43	2,48	2,52	2,55	2,59	2,61	2,63					
3,0	2,33	2,40	2,44	2,48	2,53	2,56	2,59	2,64	2,68	2,70	2,71				
3,2	2,37	2,44	2,48	2,51	2,56	2,60	2,63	2,68	2,72	2,75	2,77	2,79	2,80		
3,4	2,40	2,47	2,51	2,54	2,59	2,63	2,66	2,71	2,75	2,79	2,82	2,84	2,86	2,87	
3,6	2,43	2,49	2,53	2,56	2,61	2,65	2,68	2,74	2,78	2,81	2,85	2,87	2,90	2,91	2,93
3,8	2,45	2,51	2,55	2,58	2,63	2,67	2,70	2,75	2,80	2,83	2,87	2,90	2,92	2,95	2,97
4,0	2,47	2,53	2,57	2,60	2,65	2,68	2,73	2,77	2,81	2,85	2,88	2,91	2,94	2,97	2,99
4,2	2,49	2,54	2,58	2,61	2,66	2,69	2,73	2,78	2,82	2,86	2,89	2,92	2,95	2,99	3,01
4,4	2,50	2,56	2,59	2,62	2,67	2,70	2,73	2,78	2,83	2,86	2,90	2,93	2,96	2,99	3,02
4,6	2,51	2,57	2,60	2,63	2,68	2,71	2,74	2,79	2,83	2,87	2,90	2,94	2,97	3,00	3,03
4,8	2,52	2,58	2,61	2,64	2,68	2,72	2,75	2,80	2,84	2,87	2,91	2,94	2,97	3,00	3,03
5,0	2,53	2,58	2,62	2,64	2,69	2,72	2,75	2,80	2,84	2,88	2,91	2,95	2,97	3,00	3,03

$$l_1 = \frac{q_1 l}{K-2}; \quad l_2 = \frac{q_2 l}{K-2}; \quad l = l_1 + l = \frac{2s}{c}\sqrt{(K+1)(1-\chi)};$$
$$\lambda = \frac{n}{l}\frac{q_1^{q_1} q_2^{q_2}}{(K-2)^{K-2}}\frac{\Gamma(K)}{\Gamma(q_1+1)\,\Gamma(q_2+1)}; \quad \hat{x} = \bar{x} - \frac{s\alpha_3}{2}\frac{K+2}{K-2}; \quad \tilde{\hat{x}} = \frac{x-\hat{x}}{c};$$

n — объем выборки; c — длина интервала разбиения; l — размах кривой; $\hat{x}$ — модальное значение; $s = c \cdot \sigma$ — именованное значение среднеквадратического отклонения.

Для вычисления гамма-функции $\Gamma(z)$ приведем ряд полезных соотношений:

$$\Gamma(z) = (z-1)\Gamma(z-1) = (z-1)(z-2)\Gamma(z-2) = \ldots, \quad (z > 1);$$
$$\Gamma(z) = \frac{1}{z}\Gamma(z+1) = \frac{1}{z(z+1)}\Gamma(z+2) = \ldots, \quad (z > 1).$$

Очень полезна аппроксимация

$$\Gamma\left(1 + \frac{1}{\beta}\right) \approx 1 - 0{,}427(\beta - 1)\,\beta^{1,9},$$

при $\beta > 1$ ошибка не превышает 0,2%.

В общем случае вычисление гамма-функции по любому аргументу можно производить по формуле ($\beta > 1$, $[z]$ — целая часть числа)

$$\Gamma\left\{[z] + \frac{1}{\beta}\right\} = \left[1 - 0{,}427(\beta - 1)\,\beta^{-1,9}\right]\prod_{i=1}^{[z]-1}\left\{[z] + \frac{1}{\beta} - i\right\}.$$

Например, вычислим значение $\Gamma(5{,}6178)$:

$$\Gamma(5{,}6178) = \Gamma(5 + 0{,}6178) = \Gamma\left(5 + \frac{1}{1{,}6186}\right) =$$
$$= \left[1 - 0{,}427 \cdot 0{,}6186^{-1,9}\right] \cdot 4{,}6178 \cdot 3{,}6178 \cdot 2{,}6178 \cdot 1{,}6178 = 63{,}26699.$$

Кривая Пирсона типа I представляет собой бета-распределение (см. раздел 1.1.7), поэтому ее функция распределения $F(x; l_1, l_2, q_1, q_2, \lambda)$ может быть выражена через функцию бета-распределения

$$F(x; l_1, l_2, q_1, q_2, \lambda) = I_{\frac{l_1+x}{l_1+l_2}}(q_1 + 1, q_2 + 1)\,.$$

Задача 173. *В результате наблюдений получен статистический ряд, заданный таблицей ($n = 1000$, $c = 5$). Необходимо подобрать аппроксимирующую кривую распределения Пирсона и найти 95%-ю квантиль распределения.*

Сначала найдем моменты распределения. В нашем случае все эмпирические данные разбиты на 13 интервалов длиной $c = 5$ каждый. В качестве случайной величины будем рассматривать середину каждого интервала. Порядок вычислений представлен в таблице (здесь $\tilde{x}_i = \dfrac{x_i - 42{,}5}{5}$).

Таким образом, имеем

$$m_1 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i \cdot n_i = 0{,}093; \quad m_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i^2 \cdot n_i = 4{,}859; \quad m_3 = 7{,}755; \quad m_4 = 72{,}383.$$

i	x_i	n_i	$\tilde{x}_i$	$n_i\tilde{x}_i$	$n_i\tilde{x}_i^2$	$n_i\tilde{x}_i^3$	$n_i\tilde{x}_i^4$
1	22,5	11	−4	−44	176	−704	2816
2	27,5	93	−3	−279	837	−2511	7533
3	32,5	162	−2	−324	648	−1296	2592
4	37,5	176	−1	−176	176	−176	176
5	42,5	178	0	0	0	0	0
6	47,5	132	1	132	132	132	132
7	52,5	101	2	202	404	808	1616
8	57,5	67	3	201	603	1809	5427
9	62,5	40	4	160	640	2560	10240
10	67,5	24	5	120	600	3000	15000
11	72,5	12	6	72	432	2592	15552
12	77,5	3	7	21	147	1029	7203
13	82,5	1	8	8	64	512	4096
	$\sum$		100	93	4859	7755	72383

Далее вычисляем

$$\bar{x} = 42{,}5 + m_1 \cdot c = 42{,}5 + 0{,}093 \cdot 5 = 42{,}965; \quad \sigma^2 = m_2 - m_1^2 = 4{,}859 - 0{,}093^2 = 4{,}85;$$

$$\sigma = 2{,}202 \quad (s = c \cdot \sigma = 5 \cdot 2{,}202 = 11{,}011);$$

$$\alpha_3 = \frac{7{,}755 - 3 \cdot 4{,}859 \cdot 0{,}093 + 2 \cdot 0{,}093^2}{(4{,}850)^{\frac{3}{2}}} = 0{,}60075 \quad (\alpha_3^2 = 0{,}3609);$$

$$\alpha_4 = \frac{72{,}383 - 4 \cdot 7{,}755 \cdot 0{,}093 + 6 \cdot 4{,}859 \cdot 0{,}093^2 - 3 \cdot 0{,}093^4}{4{,}850^2} = 2{,}965.$$

Имеем

$$K = \frac{6 \cdot (2{,}965 - 0{,}3609 - 1)}{3 \cdot 0{,}3609 - 2 \cdot 2{,}965 + 6} = 8{,}3496; \quad \chi = -\frac{0{,}3609 \cdot (2 + 8{,}3496)^2}{16 \cdot (1 + 8{,}3496)} = -0{,}2584.$$

Так как $\chi < 0$, можно применить аппроксимацию эмпирического распределения кривой Пирсона типа 1.

Выполняем необходимые вычисления:

$$\hat{x} = \bar{x} - \frac{s \cdot \alpha_3}{2} \cdot \frac{K+2}{K-2} = 42{,}965 - \frac{11{,}011 \cdot 0{,}60075}{2} \cdot \frac{10{,}3496}{6{,}3496} = 37{,}574;$$

$$l = \frac{2 \cdot 11{,}011}{5} \cdot \sqrt{9{,}3496 \cdot 1{,}25841} = 15{,}107;$$

$$q_1 = \frac{1}{2} \cdot \left(6{,}3496 - 8{,}3496 \cdot \frac{0{,}60075}{4 \cdot \sqrt{9{,}3496 \cdot 1{,}25841}}\right) = 1{,}283;$$

$$q_2 = \frac{1}{2} \cdot (6{,}3496 + 3{,}7837) = 5{,}067;$$

$$l_1 = \frac{1{,}283 \cdot 15{,}107}{6{,}3496} = 3{,}052; \quad l_2 = \frac{5{,}067 \cdot 15{,}107}{6{,}3496} 12{,}055 \ (l = l_1 + l_2 = 3{,}052 + 12{,}055 = 15{,}107);$$

$$\chi = \frac{1000}{15{,}107} \cdot \frac{1{,}283^{1{,}283} \cdot 5{,}067^{5{,}067}}{6{,}3496^{6{,}3496}} \cdot \frac{\Gamma(8{,}3496)}{\Gamma(2{,}283) \cdot \Gamma(6{,}067)};$$

$$\Gamma(8{,}3496) = \Gamma\left(8 + \frac{1}{2{,}8604}\right) = \left[1 - 0{,}427 \cdot 1{,}8604 \cdot 2{,}8604^{-1{,}9}\right] \times$$
$$\times 7{,}3496 \cdot 6{,}3496 \cdot 5{,}3496 \cdot 4{,}3496 \cdot 3{,}3496 \cdot 2{,}3496 \cdot 1{,}3496 = 10289{,}8761;$$

$$\Gamma(2{,}283) = \Gamma\left(2 + \frac{1}{3{,}5335}\right) = \left[1 - 0{,}427 \cdot 2{,}5335 \cdot 3{,}5335^{-1{,}9}\right] \cdot 1{,}283 = 1{,}156879;$$

$$\Gamma(6{,}067) = \Gamma\left(6 + \frac{1}{14{,}925}\right) =$$
$$= \left[1 - 0{,}427 \cdot 13{,}925 \cdot 14{,}925^{-1{,}9}\right] \cdot 5{,}067 \cdot 4{,}067 \cdot 3{,}067 \cdot 2{,}067 \cdot 1{,}067 = 134{,}51827;$$
$$\lambda = 2{,}7135 \cdot \frac{10289{,}8761}{1{,}156879 \cdot 134{,}51827} = 179{,}419962.$$

Таким образом, уравнение для аппроксимирующей плотности распределения имеет вид

$$y = 179{,}419962 \cdot \left(1 + \frac{\tilde{\tilde{x}}}{3{,}052}\right)^{1{,}283} \cdot \left(1 - \frac{\tilde{\tilde{x}}}{12{,}055}\right)^{5{,}607}.$$

Напомним, что в качестве переменной мы используем

$$\tilde{\tilde{x}} = \frac{x - \hat{x}}{c} = 0{,}2 \cdot (x - 37{,}574).$$

Относительно реальной переменной уравнение принимает вид

$$y = 179{,}419962 \cdot (0{,}0655 \cdot x - 1{,}46225)^{1{,}283} \cdot (1{,}62337 - 0{,}01659 \cdot x)^{5{,}067}.$$

Например, для $x = 52{,}5$ частота равна

$$y(52{,}5) = 179{,}419962 \cdot (0{,}0655 \cdot 52{,}5 - 1{,}46225)^{1{,}283} \cdot (1{,}62337 - 0{,}01659 \cdot 52{,}5)^{5{,}067} = 101{,}7,$$

что совпадает с эмпирической частотой, равной 101.

Для нахождения 95%-й квантили распределения обратимся к табл. 117. Для $\alpha_3^2 = 0{,}359$, $\alpha_4 = 2{,}965$ и $\alpha = 0{,}95$ находим, что $y_{0,95} \approx 1{,}8$. Следовательно,

$$y_{0,95} = \frac{x_{0,95} - \bar{x}}{s} = \frac{x_{0,95} - 42{,}965}{11{,}011} = 1{,}8; \quad x_{0,95} = 42{,}965 + 1{,}8 \cdot 11{,}011 = 62{,}78.$$

Следовательно, вероятность того, что $x \leqslant 62{,}78$, равна 0,95.

3.6.2.2. Кривые Пирсона типа II

Для кривых этого семейства $\chi = 0$, $\alpha_3 = 0$ и $\alpha_4 < 3$. Они являются частным случаем кривых типа I и определяются уравнением

$$y = \lambda\left(1 - \frac{\tilde{x}^2}{l^2}\right)^q,$$

где

$$l = \sigma\sqrt{\frac{2\alpha_4}{3 - \alpha_4}}; \quad q = \frac{5\alpha_4 - 9}{2(3 - \alpha_4)}; \quad \lambda = \frac{n}{2^{2q+1}} \frac{\Gamma(2q+2)}{[\Gamma(q+1)]^2}; \quad \tilde{x} = \frac{x - \bar{x}}{c}.$$

Кривые симметричны относительно оси ординат и $\hat{x} = \bar{x}$.

Задача 174. *Для статистического ряда, заданного таблицей* ($n = 205$, $c = 2$) *подобрать кривую распределения из семейства Пирсона.*

Вычисление моментов представлено в таблице (здесь $\tilde{x}_i = \dfrac{x_i - 34}{2}$).

Находим ($c = 2$)

$$m_1 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i \cdot n_i = 0{,}0195; \quad m_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i^2 \cdot n_i = 2{,}1951;$$
$$m_3 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i^3 \cdot n_i = 0{,}3122; \quad m_4 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i^4 \cdot n_i = 13{,}2;$$
$$\bar{x} = 34 + 0{,}0195 \cdot 2 = 34{,}039.$$

i	x_i	n_i	$\tilde{x}_i$	$n_i\tilde{x}_i$	$n_i\tilde{x}_i^2$	$n_i\tilde{x}_i^3$	$n_i\tilde{x}_i^4$
1	26	1	−4	−4	16	−64	256
2	28	6	−3	−18	54	−162	486
3	30	27	−2	−54	108	−216	432
4	32	40	−1	−40	40	−40	40
5	34	54	0	0	0	0	0
6	36	45	1	45	45	45	45
7	38	23	2	46	92	184	368
8	40	7	3	21	63	189	567
9	42	2	4	8	32	128	512
	$\sum$	205		4	450	64	2706

В случае, когда обе ветви кривой асимптотически приближаются к оси абсцисс, необходимо при оценке четных центральных моментов ($\mu_2 = m_2 - m_1^2$ и $\mu_4 = m_4 - 4m_3 \cdot m_1 + 6m_2 \cdot m_1^2 - 3m_1^4$) применять корректирующие поправки Шеппарда [10], с учетом которых

$$\hat{\mu}_2 = \mu_2 - \frac{c^2}{12}; \quad \hat{\mu}_4 = \mu_4 - \frac{c^2}{2} \cdot \mu_2 + \frac{7c^2}{240}.$$

В нашем случае, полагая $c = 1$ (половина разряда), имеем

$$\mu_2 = 2{,}1951 - 0{,}0195^2 = 2{,}1947; \quad \hat{\mu}_2 = \sigma^2 = 2{,}1947 - 0{,}0833 = 2{,}1114 \ (\sigma = 1{,}453);$$

$$\mu_4 = 13{,}180656 - 0{,}5 \cdot 2{,}1947 + 0{,}02916 = 12{,}1125;$$

$$\hat{\mu}_4 = 13{,}2 - 4 \cdot 0{,}3122 \cdot 0{,}0195 + 6 \cdot 2{,}1951 \cdot 0{,}0195 - 3 \cdot 0{,}0195^4 = 13{,}1800656;$$

$$\mu_3 = m_3 - 3 \cdot m_2 \cdot m_1 + 2 \cdot m_1^3 = 0{,}3122 - 3 \cdot 2{,}1951 \cdot 0{,}0195 + 2 \cdot 0{,}0195^3 = 0{,}1838.$$

Тогда получаем

$$\alpha_3 = \frac{\hat{\mu}_3}{\hat{\mu}_2^{\frac{3}{2}}} = \frac{0{,}1838}{2{,}1114^{\frac{3}{2}}} = 0{,}060 \ (\alpha_3^2 = 0{,}0036); \quad \alpha_4 = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} = \frac{12{,}1125}{2{,}1114^2} = 2{,}716;$$

$$K = \frac{6 \cdot (2{,}716 - 0{,}0036 - 1)}{3 \cdot 0{,}0036 - 2 \cdot 2{,}716 + 6} = 17{,}751; \quad \chi = \frac{0{,}0036 \cdot 19{,}751^2}{16 \cdot 18{,}751} = -0{,}00468.$$

Так как $\alpha_4 > 3$ и $\chi \approx 0$, для аппроксимации заданного статистического ряда можно использовать кривую типа II.

Находим параметры кривой:

$$q = \frac{5 \cdot 2{,}716}{2(3 - 2{,}716)} = 8{,}063; \quad l = 1{,}453 \cdot \sqrt{\frac{2 \cdot 2{,}716}{3 - 2{,}716}} = 6{,}354;$$

$$\lambda = \frac{205}{2^{17{,}126} \cdot 6{,}354} \cdot \frac{\Gamma(18{,}126)}{[\Gamma(9{,}063)]^2} = 225{,}56210^{-6} \cdot \frac{\Gamma(18{,}126)}{[\Gamma(9{,}063)]^2}.$$

Далее вычисляем

$$\Gamma(18{,}126) = \Gamma\left(18 + \frac{1}{7{,}9365}\right) = \left[1 - 0{,}427 \cdot 6{,}9365 \cdot 7{,}9365^{-1{,}9}\right] \cdot 17{,}126 \ldots 1{,}126 = 5{,}108028 \cdot 10^4;$$

$$\Gamma(9{,}063) = \Gamma\left(9 + \frac{1}{15{,}873}\right) = \left[1 - 0{,}427 \cdot 14{,}873 \cdot 15{,}873^{-1{,}9}\right] \cdot 8{,}063 \ldots 1{,}063 = 46124{,}251.$$

Окончательно

$$\lambda = 225{,}562 \cdot 10^{-6} \cdot \frac{5{,}108028 \cdot 10^{14}}{46124{,}251^2} = 54{,}15774468.$$

Таким образом, уравнение аппроксимирующей кривой имеет вид

$$y = 54{,}15774468 \cdot \left(1 - \frac{\tilde{x}^2}{40{,}373316}\right)^{8{,}063}.$$

Переменной здесь является $\tilde{x} = \dfrac{x - \bar{x}}{c} = \dfrac{x - 34{,}039}{2}$.

Окончательно имеем

$$y = 54{,}15774468 \cdot (0{,}421553186 \cdot x - 0{,}0061922 \cdot x^2 - 6{,}174624)^{8{,}063}.$$

Например, для $x = 30$:

$$y(30) = 54{,}15774468 \cdot \left(0{,}421553186 \cdot 30 - 0{,}0061922 \cdot 30^2 - 6{,}174624\right)^{8{,}063} = 23,$$

что близко к эмпирической частоте, равной 27.

3.6.2.3. Кривые Пирсона типа III

Для этого типа кривых $\chi = \pm\infty$. Уравнение кривой имеет вид

$$y = \lambda \left(1 + \frac{\tilde{\tilde{x}}}{l}\right)^q \exp\left(-\frac{q\tilde{\tilde{x}}}{l}\right),$$

где

$$q = \frac{4}{\alpha_3^2} - 1; \quad l = \sigma\left(\frac{2}{\alpha_3} - \frac{\alpha_3}{2}\right); \quad \tilde{\tilde{x}} = \frac{x - \hat{x}}{c}; \quad \hat{x} = \bar{x} - \frac{c\sigma\alpha_3}{2};$$

$$\lambda = \frac{n}{l}(q+2)\left[\exp(q+1)\,\Gamma(q+2)\right]^{-1}.$$

Кривые типа III асимметричны и ограничены в одном направлении точкой $x = -l$. Кривая типа III совпадает с гамма-распределением. Хорошая аппроксимация эмпирических распределений кривой типа III достигается уже при $|\chi| > 4$.

Задача 175. *Найти аппроксимирующую кривую для плотности распределения случайной величины, заданной эмпирическим статистическим рядом* ($n = 40000$, $c = 2$), *приведенным в таблице.*

Вычисления начальных моментов приведены в таблице (здесь $\tilde{x}_i = \dfrac{x_i - 2{,}356}{2}$):

i	x_i	n_i	$\tilde{x}_i$	$n_i\tilde{x}_i$	$n_i\tilde{x}_i^2$	$n_i\tilde{x}_i^3$	$n_i\tilde{x}_i^4$
1	1,5	9600	−0,448	−19200	38400	−76800	153600
2	3,5	13600	0,552	−13600	13600	−13600	13600
3	5,5	9200	1,552	0	0	0	0
4	7,5	4000	2,552	4000	4000	4000	4000
5	9,5	2000	3,552	4000	8000	16000	32000
6	11,5	800	4,552	2400	7200	21600	64800
7	13,5	400	5,552	1600	6400	25600	102400
8	15,5	0	6,552	0	0	0	0
9	17,5	400	7,552	2400	14400	86400	518400
	$\sum$	40000		−18400	92000	63200	888800

Вычисляем

$$m_1 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i \cdot n_i = -0{,}46; \quad m_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i^2 \cdot n_i = 2{,}30;$$

$$m_3 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i^3 \cdot n_i = 1{,}58; \quad m_4 = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i^4 \cdot n_i = 22{,}22;$$

$$\sigma^2 = \mu_2 = m_2 - m_1^2 = 2{,}0884; \quad \sigma = 1{,}4451;$$

$$\mu_3 = 1{,}58 - 3 \cdot 2{,}30 \cdot (-0{,}46) + 2 \cdot (-0{,}46)^3 = 4{,}5593;$$

$$\mu_4 = 22{,}22 - 4 \cdot 1{,}58 \cdot (-0{,}46) + 6 \cdot 2{,}30 \cdot (-0{,}46)^2 - 3 \cdot 0{,}46^4 = 27{,}9129;$$

$$\bar{x} = 5{,}5 - 0{,}46 \cdot 2 = 4{,}58; \quad \alpha_3 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = 1{,}1508; \quad \alpha_3^2 = 2{,}2825;$$

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{27{,}9129}{2{,}0884^2} = 6{,}400; \quad K = \frac{6 \cdot (6{,}4 - 2{,}2825 - 1)}{3 \cdot 2{,}2825 - 2 \cdot 6{,}4 + 6} = 393{,}79; \quad \chi = -56{,}6.$$

Так как $|\chi| \gg 4$, то приемлема аппроксимация кривой Пирсона типа III. Вычисляем параметры кривой:

$$q = \frac{4}{\alpha_3^2} - 1 = \frac{4}{2{,}2822} - 1 = 0{,}752451; \quad l = 1{,}4451 \cdot \left(\frac{2}{1{,}5108} - \frac{1{,}5108}{2}\right) = 0{,}8213977;$$

$$\hat{x} = 4{,}58 - \frac{2 \cdot 1{,}4451 \cdot 1{,}5108}{2} = 2{,}3974292;$$

$$\lambda = \frac{40000}{0{,}8213977} \cdot \frac{2{,}752451}{\exp(1{,}752451) \cdot \Gamma(2{,}752451)} = \frac{232356451}{\Gamma(2{,}752451)}.$$

Находим

$$\Gamma(2{,}752451) = \Gamma\left(2 + \frac{1}{1{,}32899}\right) = \left[1 - 0{,}427 \cdot 0{,}32899 \cdot 1{,}32899^{-1{,}9}\right] \cdot 1{,}752451 = 1{,}6090457;$$

$$\lambda = \frac{23235{,}19417}{1{,}6090457} = 14440{,}3569.$$

Уравнение кривой имеет вид

$$y = 14440{,}3589 \cdot \left(1 + \frac{\tilde{\tilde{x}}}{0{,}8213977}\right)^{0{,}752451} \cdot \exp\left(-\frac{0{,}752451 \cdot \tilde{\tilde{x}}}{0{,}5213977}\right).$$

Здесь в качестве переменной используется величина $\tilde{\tilde{x}} = \dfrac{x - \hat{x}}{c} = 0{,}5x - 1{,}1925$, и относительно нее после преобразований получаем

$$y = 43052{,}93758 \cdot (0{,}608718529 \cdot x - 0{,}451793692)^{0{,}752451} \cdot \exp(-0{,}458030866 \cdot x).$$

Например, для $x = 3{,}5$ имеем

$$y = 43052{,}93758 \cdot (0{,}608718529 \cdot 3{,}5 - 0{,}451793692)^{0{,}752451} \cdot \exp(-0{,}458030866 \cdot 3{,}5) = 12796,$$

что близко к эмпирической частоте, равной 13600.

3.6.2.4. Кривые Пирсона типа IV

Для кривых этого семейства имеет место $0 < \chi < 1$. Уравнение кривой типа IV имеет вид

$$y = \lambda \left(1 + \frac{\tilde{\tilde{x}}^2}{l^2}\right)^{-q} \exp\left(-\nu \operatorname{arctg} \frac{\tilde{\tilde{x}}}{l}\right),$$

где

$$q = \frac{2 + r}{2}; \quad r = \frac{6(\alpha_4 - \alpha_3^2 - 1)}{2\alpha_4 - 3\alpha_3^2 + 6}; \quad \nu = -\frac{r(r-2)\,\alpha_3}{\sqrt{16(r-1) - \alpha_3^2(r-2)^2}};$$

$$l = \frac{\sigma}{4}\sqrt{16(r-1) - \alpha_3^2(r-2)^2}; \quad \lambda = \frac{n}{l}\frac{1}{F(r, \nu)};$$

$F(r, \nu)$ — функция, значения которой по аргументу $\varphi = \operatorname{arctg} \dfrac{|\nu|}{r}$ приведены в [10].

Уравнение записано относительно переменной $\tilde{\tilde{x}} = \dfrac{x - \tilde{x}}{c} - \dfrac{\nu}{r}l$.

Кривые асимметричны и имеют неограниченный размах.

Задача 176. *Для эмпирического ряда* ($N = 1000$, $c = 2$), *приведенного в таблице, найти аппроксимирующую кривую из семейства Пирсона.*

Вычисление начальных моментов приведено в таблице ($\tilde{x}_i = \dfrac{x_i - 1{,}9}{0{,}2}$):

i	x_i	n_i	$\tilde{x}_i$	$n_i\tilde{x}_i$	$n_i\tilde{x}_i^2$	$n_i\tilde{x}_i^3$	$n_i\tilde{x}_i^4$
1	0,5	−7	2	−14	98	−686	4802
2	0,7	6	5	−30	180	−1080	6480
3	0,9	−5	14	−70	350	−1750	8750
4	1,1	−4	33	−132	528	−2112	8448
5	1,3	−3	67	−201	603	−1809	5427
6	1,5	−2	123	−246	492	−984	1968
7	1,7	−1	160	−160	160	−160	160
8	1,9	0	172	0	0	0	0
9	2,1	1	153	153	153	153	153
10	2,3	2	112	224	448	896	1792
11	2,5	3	68	204	612	1836	5508
12	2,7	4	41	164	656	2624	10496
13	2,9	5	25	125	625	3125	15625
14	3,1	6	14	84	504	3024	18144
15	3,3	7	7	49	343	2401	16807
16	3,5	8	3	24	192	1536	12288
17	3,7	9	1	9	81	729	6561
	$\sum$	1000		183	6025	7743	123409

Имеем

$$m_{1n} = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i \cdot n_i = 0{,}183; \quad m_2 = 6{,}025; \quad m_3 = 7{,}743; \quad m_4 = 123{,}409;$$

$$\hat{\mu}_2 = m_2 - m_1^2 - \frac{1}{12} = 5{,}908; \quad \mu_3 = m_3 - 3 \cdot m_2 \cdot m_1 + 2m_1^3 = 4{,}448;$$

$$\mu_4 = m_4 - 4 \cdot m_3 \cdot m_1 + 6 \cdot m_2 \cdot m_1^2 - 3 \cdot m_1^4 = 118{,}948;$$

$$\hat{\mu}_4 = \mu_4 - \frac{\hat{\mu}_2}{12} + \frac{7}{240} = 118{,}484; \quad \sigma = \sqrt{\hat{\mu}_2} = 2{,}43064;$$

$$\alpha_3 = \frac{\mu_3}{\hat{\mu}_2^{\frac{3}{2}}} = 0{,}30974; \quad \alpha_3^2 = 0{,}096; \quad \alpha_4 = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} = 3{,}407816.$$

Вычисляем далее:

$$K = \frac{6 \cdot (3{,}407816 - 0{,}096 - 1)}{3 \cdot 0{,}096 - 2 \cdot 3{,}407816 + 6} = -26{,}288959; \quad \chi = -\frac{\alpha_3^2 \cdot (K+1)^2}{16 \cdot (K+1)} = -0{,}139971.$$

Таким образом, для аппроксимации может быть выбрана кривая типа IV из семейства Пирсона. Вычисляем ее параметры:

$$r = -K = 26{,}288959; \quad q = \frac{2 + 26{,}288959}{2} = 14{,}14444795;$$

$$l = \frac{2{,}43064}{4} \cdot \sqrt{16 \cdot 25{,}288959 - 0{,}0961 \cdot 24{,}288959^2} = 11{,}3345908;$$

$$\nu = -\frac{26{,}288959 \cdot 24{,}288959 \cdot 0{,}30974}{\sqrt{16 \cdot 25{,}288959 - 0{,}096 \cdot 24{,}288959^2}} = -10{,}602238; \quad \lambda = \frac{n}{l} \cdot \frac{1}{F(r,\nu)} = \frac{88{,}225505}{F(r,\nu)}.$$

Находим из [10, таблица XIV, с. 532]: $F(r,\nu) \approx 1{,}82$ для

$$\varphi = \operatorname{arctg} \frac{|\nu|}{r} = \operatorname{arctg} \frac{11{,}12625}{226{,}576} = 22{,}6^\circ; \quad \lambda = \frac{88{,}225505}{1{,}82} = 48{,}4753.$$

Далее $\tilde{\tilde{x}} = \dfrac{x-1{,}9}{0{,}2} + \dfrac{10{,}602238\cdot 11{,}3345908}{26{,}225505} = 5x - 4{,}928023.$

Окончательно получаем уравнение кривой

$$y = 48{,}4753\cdot\left(1+\frac{\tilde{\tilde{x}}}{4{,}3345908^2}\right)^{-14{,}1444795}\cdot\exp\left(10{,}602238\,\mathrm{arctg}\,\frac{\tilde{\tilde{x}}}{11{,}3345908}\right) =$$

$$= 48{,}4753\cdot\left[1+\frac{(5x-4{,}9288023)}{11{,}335908^2}\right]^{-14{,}1444795}\cdot\exp\left[10{,}602238\cdot\mathrm{arctg}\left(\frac{5x-4{,}9288023}{4{,}3345908}\right)\right].$$

3.6.2.5. Кривые Пирсона типа V

Для кривых этого типа $\chi = 1$. Уравнение кривой типа V имеет вид

$$y = \lambda\tilde{\tilde{x}}^{-q}\exp\left\{-\frac{\nu}{\tilde{\tilde{x}}}\right\},$$

где

$$q = 4+\frac{8+4\sqrt{4+\alpha_3^2}}{\alpha_3^2};\quad \nu = \sigma(q-2)\sqrt{q-3}\ (\text{знак такой же, как у } \alpha_3);$$

$$\lambda = n\frac{\nu^{q-1}}{\Gamma(q-1)};\quad \tilde{\tilde{x}} = \frac{x-\bar{x}}{c}+\frac{\nu}{q-2}.$$

Кривые симметричны и определены для $0 < x < \infty$.

Задача 177. *Для статистического ряда, приведенного в таблице, найти аппроксимирующую кривую из семейства Пирсона* ($n = 1000$, $c = 0{,}04$).

Имеем $\tilde{x}_i = \dfrac{x_i - 0{,}2}{0{,}04}$ и сводим результаты вычислений в таблицу:

i	x_i	n_i	$\tilde{x}_i$	$n_i\tilde{x}_i$	$n_i\tilde{x}_i^2$	$n_i\tilde{x}_i^3$	$n_i\tilde{x}_i^4$
1	0,04	−4	3	−12	48	−192	768
2	0,08	−3	11	−33	99	−297	891
3	0,12	−2	99	−198	396	−792	1584
4	0,16	−1	346	−346	346	−346	346
5	0,2	0	288	0	0	0	0
6	0,24	1	148	148	148	148	148
7	0,28	2	83	168	332	664	1328
8	0,32	3	15	45	135	405	1215
9	0,36	4	6	24	96	384	1536
10	040	5	1	5	25	125	625
		$\sum$	1000	−199	1625	99	8441

Находим:

$$m_1 = -0{,}199;\quad m_2 = 1{,}625;\quad m_3 = 0{,}099;\quad m_4 = 8{,}441;$$

$$\mu_2 = m_2 - m_1^2 = 1{,}585;\quad \mu_3 = m_3 - 3m_2m_1 + 2m_1^3 = 1{,}053;$$

$$\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4 = 8{,}901;\quad \sigma = \sqrt{\mu_2} = 1{,}259;$$

$$\bar{x} = 0{,}20 + c\cdot m_1 = 0{,}192;\quad \alpha_3 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = 0{,}528;\quad \alpha_3^2 = 0{,}278;\quad \alpha_4 = \frac{\mu_4}{\mu_2^2} = 3{,}549.$$

Далее $K = \dfrac{6\cdot(3{,}543-0{,}278-1)}{3\cdot 0{,}278 - 2\cdot 3{,}543+6} = -53{,}928$; $\chi = 0{,}885$.

Так как $\chi \approx 1$, для аппроксимации может быть использована кривая из семейства Пирсона типа V.

Вычисляем параметры кривой

$$q = 4\frac{8 + 4\cdot\sqrt{4 + 0{,}278}}{0{,}278} = 62{,}537; \qquad \nu = 1{,}259 \cdot 60{,}637 \cdot \sqrt{59{,}537} = 588{,}085;$$

$$\lambda = 1000 \cdot \frac{588{,}085^{61{,}537}}{\Gamma(61{,}537)} = 3{,}5514 \cdot 10^{90}.$$

Уравнение кривой принимает вид

$$y = 3{,}5514 \cdot 10^{90} \cdot \tilde{\tilde{x}}^{-62{,}537} \cdot \exp\left(-\frac{588{,}085}{\tilde{\tilde{x}}}\right).$$

В нашем случае переменная $\tilde{\tilde{x}} = \dfrac{x - 0{,}192}{0{,}04} + \dfrac{588{,}085}{60{,}537} = 25x + 4{,}9145$ и

$$y = 3{,}5514 \cdot 10^{90} \cdot (25x + 4{,}9145)^{-62{,}537} \cdot \exp\left(-\frac{588{,}085}{25x + 4{,}9145}\right).$$

Например, для $x = 0{,}16$ имеем

$$y = 3{,}5514 \cdot 10^{90} \cdot 8{,}91456^{-62{,}537} \cdot \exp\left(-\frac{588{,}085}{8{,}91456}\right) = 304,$$

что близко к эмпирическому значению 346.

3.6.2.6. Кривые Пирсона типа VI

Для кривых этого типа справедливо неравенство $1 < \chi < \infty$.

Уравнение кривой имеет вид

$$y = \lambda \tilde{\tilde{x}}^{-q_1} \left(\tilde{\tilde{x}} - l\right)^{q_2},$$

где

$$\frac{q_1}{q_2} = \frac{1}{2}\left\{K(K+2)\frac{\alpha_3}{\sqrt{\alpha_3^2(K+2)^2 + 16(K+1)}} \mp (K-2)\right\};$$

$$l = 2\sigma\sqrt{(K+1)(1-\chi)}; \quad \lambda = n\frac{l^{q_1 - q_2 - 1}\Gamma(q_1)}{\Gamma(q_1 - q_2 - 2)}; \quad \tilde{\tilde{x}} = \frac{x - \hat{x}}{c} + \frac{l(q_1 - 1)}{q_1 - q_2 - 2}.$$

Кривые асимметричны и ограничены в одном направлении точкой $x = l$. Если $\alpha_3 < 0$, то $l < 0$ и размах распределения находится в пределах $(\infty, -1)$.

Задача 178. *Для статистического ряда, приведенного в таблице, найти аппроксимирующую кривую из семейства Пирсона и 99%-ю квантиль распределения* ($n = 368$, $c = 1$).

Здесь $x_i = x - 3$.

i	x_i	n_i	$\tilde{x}_i$	$n_i\tilde{x}_i$	$n_i\tilde{x}_i^2$	$n_i\tilde{x}_i^3$	$n_i\tilde{x}_i^4$
1	1	−2	1	−2	4	−8	16
2	2	−1	56	−56	56	−56	56
3	3	0	167	0	0	0	0
4	4	1	98	98	98	98	98
5	5	2	34	68	136	272	544
6	6	3	9	27	81	243	729
7	7	4	2	8	32	128	512
8	8	5	1	5	25	125	625
		$\sum$	368	148	432	802	2580

Находим

$$m_1 = 0{,}4021; \quad m_2 = 1{,}17391; \quad m_3 = 2{,}17935; \quad m_4 = 7{,}01087.$$

Учитывая, что размах кривой неограничен, при оценке центральных моментов будем учитывать поправки Шеппарда ($c = 1$). Имеем

$$\mu_2 = m_2 - m_1^2 = 1{,}0121; \quad \hat{\mu}_2 - \mu_2 - \frac{1}{12} = 0{,}92883;$$

$$\mu_3 = 2{,}17935 - 3 \cdot 1{,}17391 \cdot 0{,}40217 + 2 \cdot 0{,}40217^2 = 0{,}89310;$$

$$\mu_4 = 4{,}56573; \quad \hat{\mu}_4 = \mu_4 - \frac{\hat{\mu}_2}{2} + \frac{7}{240} = 4{,}088; \quad \sigma = \sqrt{\hat{\mu}_2} = 0{,}96376;$$

$$\alpha_3 = \frac{\mu_3}{\hat{\mu}_2^{\frac{3}{2}}} = 0{,}9977; \quad \alpha_4 = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} = 4{,}7394; \quad \bar{x} = 3 + 0{,}40217 \cdot 1 = 3{,}40217.$$

Далее

$$K = \frac{6 \cdot (4{,}7394 - 0{,}9954 - 1)}{3 \cdot 0{,}9954 - 2 \cdot 4{,}7394 + 6} = -33{,}423; \quad \chi = -\frac{0{,}9954 \cdot (-31{,}423)^2}{16 \cdot (-32{,}423)} = 1{,}761.$$

Видим, что возможна аппроксимация кривой типа VI. Находим параметры кривой:

$$q_1 = \frac{1}{2} \cdot \{(-33{,}423) \cdot (-31{,}423)\} \cdot \frac{0{,}9977}{\sqrt{0{,}9954 \cdot (-31{,}423) + 16 \cdot (-32{,}423)}} - (-35{,}423) = 42{,}031;$$

$$q_2 = 6{,}608; \quad l = 2 \cdot 0{,}96376 \cdot \sqrt{(-32{,}423) \cdot (1 - 1{,}761)}9{,}5745;$$

$$\lambda = 368 \cdot \frac{9{,}5745^{42{,}031 - 6{,}608 - 1} \cdot \Gamma(42{,}031)}{\Gamma(34{,}423) \cdot \Gamma(7{,}608)} = 1{,}5268527 \cdot 10^{46};$$

$$\tilde{\tilde{x}} = \frac{x - \hat{x}}{c} + 12{,}345 = \frac{x - 3}{1} + 12{,}345 = x + 9{,}345.$$

Окончательно получаем

$$y = 1{,}5268627 \cdot 10^{46} \cdot (x + 9{,}345)^{-42{,}031} \cdot (x - 1{,}0383)^{6{,}608}.$$

Например, при $x = 3$ имеем $y = 174$, что близко к эмпирическому значению 167.

Для нахождения 99%-й квантили обратимся к табл. 117. Из нее следует для $\alpha_3^2 = 0{,}9954$ и $\alpha_4 = 4{,}74$, что $y_{0,99} = 3{,}03$ и

$$x_{0,99} = \bar{x} + \sigma \cdot c \cdot y_{0,99} = 3{,}40217 + 0{,}96376 \cdot 1 \cdot 3{,}03 = 6{,}32236.$$

3.6.2.7. Кривые Пирсона типа VII

Для этого случая $\chi = 0$, $\alpha_3 = 0$ и $\alpha_4 > 3$. Уравнение кривой имеет вид

$$y = \lambda \left(1 + \frac{\tilde{x}^2}{l^2}\right)^{-q},$$

где

$$q = \frac{5\alpha_4 - 9}{2(\alpha_4 - 3)} > 0; \quad l = \sigma \sqrt{\frac{2\alpha_4}{\alpha_4 - 3}}; \quad \lambda = \frac{n}{l} \frac{\Gamma(q)}{\Gamma\left(q - \frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}; \quad \tilde{x} = \frac{x - \bar{x}}{c}.$$

Кривые асимметричны относительно среднего значения, совпадающего с модой, и имеют неограниченный размах. При $\chi = 0$, $\alpha_3 = 0$ и $\alpha_4 = 3$ распределение переходит в нормальное.

Задача 179. *Для статистического ряда, заданного таблицей, найти аппроксимирующее распределение из семейства Пирсона* ($n = 2886$, $c = 10$).

Здесь $x_{0,99} = \bar{x} + \sigma \cdot c \cdot y_{0,99} = 3{,}40217 + 0{,}96376 \cdot 1 \cdot 3{,}03 = 6{,}6322$.

Результаты расчетов сведем в таблицу:

i	x_i	n_i	$\tilde{x}_i$	$n_i\tilde{x}_i$	$n_i\tilde{x}_i^2$	$n_i\tilde{x}_i^3$	$n_i\tilde{x}_i^4$
1	725	−5	4	−20	100	−500	2500
2	735	−4	10	−40	160	−640	2560
3	745	−3	84	−252	756	−2268	6804
4	755	−2	270	−540	1080	−2160	4320
5	765	−1	658	−658	658	−658	658
6	775	0	1095	0	0	0	0
7	785	1	506	506	506	506	506
8	795	2	152	304	608	1216	2432
9	805	3	63	189	567	1701	5103
10	815	4	32	128	512	2048	8192
11	825	5	12	60	300	1500	7500
		$\sum$	2886	−323	5247	745	40575

Далее:

$$q = \frac{5 \cdot 4{,}390574 - 9}{2 \cdot (4{,}390574 - 3)} = 4{,}657382; \quad l = 1{,}34365 \cdot \sqrt{\frac{2 \cdot 4{,}390574}{1{,}390574}} = 3{,}376484;$$

$$\lambda = \frac{2886}{3{,}376484} \cdot \frac{\Gamma(4{,}657382)}{\Gamma(4{,}657382) \cdot \Gamma(0{,}5)};$$

$$\Gamma(4{,}657382) = \left[1 - 0{,}427 \cdot 0{,}521185 \cdot 1{,}521185^{-1{,}9}\right] \cdot 3{,}657382 \ldots 1{,}657382 = 14{,}492645;$$

$$\Gamma(4{,}157382) = \left[1 - 0{,}427 \cdot 5{,}353967 \cdot 6{,}353967^{-1{,}9}\right] \cdot 3{,}157382 \ldots 1{,}157382 = 7{,}3466223;$$

$$\Gamma(0{,}5) = \frac{\Gamma(1{,}5)}{0{,}5} = \frac{1 - 0{,}427 \cdot 1 \cdot 2^{-1{,}9}}{0{,}5} = 1{,}771176;$$

$$\lambda = \frac{2886}{3{,}376484} \cdot \frac{14{,}492645}{7{,}3466223 \cdot 1{,}771176} = 951{,}98444575.$$

Имеем уравнение искомой кривой

$$y = 951{,}9844785 \cdot \left(1 + \frac{\tilde{x}^2}{3{,}376484^2}\right)^{-4{,}657382}.$$

В нашем случае переменной является

$$\tilde{x} = \frac{x - \bar{x}}{c} = 0{,}1 \cdot (x - \bar{x})\,; \quad \bar{x} = 775 + m_1 \cdot c = 775 + 10 \cdot (-0{,}112) = 773{,}88;$$

$$\tilde{x} = 0{,}1x - 77{,}388.$$

Относительно реальной переменной уравнение кривой имеет вид

$$y = 951{,}984475 \cdot \left(1 + \frac{(0{,}1x - 77{,}388)^2}{3{,}376484^2}\right)^{-4{,}657382}.$$

Например, при $x = 795$ имеем

$$y(795) = 951{,}984475 \cdot \left(0{,}000877145 \cdot 795^2 - 1{,}3576075 \cdot 795 + 526{,}3126435\right)^{-4{,}657382} = 204,$$

что находится вблизи эмпирического значения 152.

Внимание! Во всех расчетах кривых Пирсона требуется высокая точность вычислений (необходимо удерживать до $8 \div 10$ знаков после запятой), что объясняется мультипликативной схемой накопления ошибок в степенных членах.

3.6.3. Разложение теоретических распределений

Для эмпирического распределения можно получить хорошее аналитическое приближение, если использовать плотность известного теоретического распределения $\varphi(x)$ и ее производные.

Для распределений, незначительно отличающихся от нормального, хорошие результаты дает приближение с помощью рядов Грама–Шарлье [10]

$$f(x) = \varphi(x) + \frac{\alpha_3}{6}\,\varphi^{\mathrm{III}}(x) + \frac{\alpha_4 - 3}{24}\,\varphi^{\mathrm{IV}}(x),$$

где $\varphi(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$ — плотность стандартного нормального распределения; $\varphi^{\mathrm{III}}(x) = -(x^3 - 3x)\varphi(x)$ — третья производная от $\varphi(x)$; $\varphi^{\mathrm{IV}}(x) = (x^4 - 6x^2 + 3)\varphi(x)$ — четвертая производная от $\varphi(x)$.

Переменная используется в нормализованной форме

$$\tilde{x} = \frac{x - \mathbf{M}(x)}{\sqrt{\mathbf{D}(x)}} = \frac{x - \bar{x}}{s}.$$

Задача 180. *Для статистического ряда, заданного таблицей, найти аппроксимирующую кривую для плотности распределения с помощью разложения нормальной плотности распределения вероятностей* ($n = 2000$, $c = 40$).

Вычисление моментов распределения приведено в таблице ($\bar{x}_i = \dfrac{x_i - 415}{40}$):

i	x_i	n_i	$\tilde{x}_i$	$n_i\tilde{x}_i$	$n_i\tilde{x}_i^2$	$n_i\tilde{x}_i^3$	$n_i\tilde{x}_i^4$
1	175	−6	0	0	0	0	0
2	215	−5	7	−35	175	−875	4375
3	255	−4	22	−88	352	−1408	5632
4	295	−3	102	−306	918	−2754	8262
5	335	−2	260	−520	1040	−2080	4160
6	375	−1	386	−386	386	−386	386
7	415	0	461	0	0	0	0
8	455	1	356	356	356	356	356
9	495	2	239	478	956	1912	3824
10	535	3	108	324	972	2916	8748
11	575	4	40	160	640	2560	10240
12	615	5	15	75	375	1875	9375
13	655	6	4	24	144	864	5184
		$\sum$	2000	82	6314	2980	60542

Имеем

$$m_1 = 0{,}041; \quad m_2 = 9{,}157; \quad m_3 = 1{,}49; \quad m_4 = 30{,}271; \quad \bar{x} = 415 + 40 \cdot 0{,}041 = 416{,}64;$$

$$\mu_2 = m_{-} m_1^2 = 3{,}1553; \quad \sigma = 1{,}77632; \quad s = \sigma \cdot c = 71{,}0528;$$

$$\mu_3 = m - 3m_2 m + 2m_1^3 = 1{,}1018; \quad \mu_4 = m_4 - 4m_3 m_1 + 6m_2 m_1^2 - 3m_1^4 = 30{,}269;$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = 0{,}1966; \quad \alpha_4 = \frac{\mu_4}{\sigma^4} = 3{,}04.$$

Для аппроксимирующей кривой по переменной $\tilde{x} = \dfrac{x - 416{,}64}{71{,}0528}$ имеем

$$f(x) = \varphi(x) \cdot \left[1 + \frac{\alpha_3}{6} \cdot \left(x^3 - 3x\right) + \frac{\alpha_4 - 3}{24} \cdot \left(x^4 - 6x^2 + 3\right)\right] =$$

$$= 0{,}3989 \cdot e^{-\frac{x^2}{2}} \cdot \left[1 + 0{,}03276 \cdot \left(x^3 - 3x\right) + 1{,}667 \cdot 10^{-3} \cdot \left(x^4 - 6x^2 + 3\right)\right].$$

Например, для $x = 495$

$$\tilde{x} = \frac{495 - 416{,}64}{71{,}0528} = 1{,}1028; \quad e^{-\frac{x^2}{2}} = 0{,}54439; \quad x^3 - 3x = -1{,}96721;$$

$$x^4 - 6x^2 + 3 = -2{,}81794;$$

$$f(1{,}1028) = 0{,}3989 \cdot 0{,}54439 \cdot \left[1 + 0{,}03276 \cdot (-1{,}96721) + 1{,}667 \cdot 10^3 \cdot (-2{,}81794)\right] = 0{,}20214.$$

Частота в соответствующем интервале равна

$$n \cdot f(x) = 2000 \cdot 0{,}20214 = 404.$$

3.6.4. Метод вкладов

Рассмотренные ранее в разделах 6.1, 6.2 и 6.3 методы восстановления функции распределения вероятностей требуют большого количества данных и неприменимы для малых выборок [370, 371]. Поэтому, начиная с 50-х годов, ведутся работы по поиску эффективных методов построения функций распределения вероятностей случайной величины по информации, содержащейся в малой выборке [371–373, 382–384]. Среди таких методов можно отметить метод вкладов с использованием бета-распределения [374, 375], метод структурной минимизации риска [377], метод уменьшения неопределенности [379], метод нормальных вкладов [380], метод последовательных медиан [381].

Широкое распространение получил метод вкладов. Суть метода вкладов заключается в том, что каждому выборочному значению случайной величины x_i ставится в соответствие некоторая непрерывная функция $\varphi_i(x_j)$, называемая функцией вклада. Наибольшую эффективность демонстрирует метод, в котором в качестве вкладов используется бета-функция [374]

$$\varphi_i(x_j) = c_i(x_j - a)^{n_i - 1}(b - x_j)^{m_i - 1}, \quad \text{где} \quad c_i = \frac{\Gamma(n_i + m_i)}{(b - a)^{n_i + m_i - 1}\Gamma(n_i)\Gamma(m_i)};$$

$a < x < b$ — область существования случайной величины; x_j — реализация случайной величины, совпадающая с модой частного бета-распределения $\varphi_i(x_j)$.

Численным экспериментом с использованием оптимизирующих критериев [376, 378] получены следующие оценки параметров частных бета-вкладов

$$n_1 = 1 + \frac{5{,}7n^{0{,}5254}}{\dfrac{b - x_i}{x_i - a} + 1}; \quad m_i = 1 + 5{,}7n^{0{,}5254} - n_i;$$

$$a = x_1 - \frac{0{,}3996n^{-0{,}2335}\,(x_n - x)}{1 - 0{,}6286n^{-0{,}2380} - 0{,}3996n^{-0{,}2335}};$$

$$b = x_1 + \frac{\left(1 - 0{,}3996n^{-0{,}2335}\right)(x_n - x_1)}{1 - 0{,}6286n^{-0{,}2380} - 0{,}3996n^{-0{,}2335}},$$

где n — объем выборки; x_1, x_n — соответственно наименьшее и наибольшее выборочные значения.

Итоговая плотность распределения в точке x_j подсчитывается по формуле $f(x_j) = \dfrac{1}{n}\sum\limits_{i=1}^{n}\varphi_i(x_j)$.

Задача 181. *В результате эксперимента получено пять значений случайной величины x_j: 1, 2, 4, 7 и 10. Необходимо с помощью метода бета-вкладов построить оценку плотности распределения вероятностей случайной величины x.*

Вычисляем

$$a = 1 - \frac{0{,}3996 \cdot 5^{-0{,}2335} \cdot (10-1)}{1 - 0{,}6286 \cdot 5^{-0{,}2380} - 0{,}3996 \cdot 5^{-0{,}2335}} = -7{,}315423;$$

$$b = 1 + \frac{\left(1 - 0{,}3996 \cdot 5^{0{,}2335}\right) \cdot (10-1)}{1 - 0{,}6286 \cdot 5^{-0{,}23380} - 0{,}3996 \cdot 5^{-0{,}2335}} = 22{,}98637;$$

$$n_\imath = 1 + \frac{5{,}7 \cdot 5^{0{,}5254}}{\dfrac{22{,}98637 - x_\imath}{x_\imath + 7{,}315423} + 1}; \quad m_i = 15{,}2774 - n_\imath.$$

Вычисляем параметры бета-вкладов:

$$n_1 = 1 + \frac{5{,}7 \cdot 5^{0{,}5254}}{\dfrac{22{,}98637 - 1}{1 + 7{,}315423} + 1} = 4{,}643588; \quad m_1 = 15{,}2774 - 4{,}643588 = 10{,}633812;$$

$$n_2 = 1 + \frac{13{,}2774}{\dfrac{22{,}98637 - 21}{2 + 7{,}315423} + 1} = 5{,}08176; \quad m_2 = 10{,}195639;$$

$$n_3 = 1 + \frac{13{,}2774}{\dfrac{22{,}98637 - 4}{4 + 7{,}315423} + 1} = 5{,}9581; \quad m_3 = 9{,}31929;$$

$$n_4 = 1 + \frac{13{,}2774}{\dfrac{22{,}98637 - 7}{7 + 7{,}315423} + 1} = 7{,}2726; \quad m_4 = 8{,}0048;$$

$$n_5 = 1 + \frac{13{,}2774}{\dfrac{22{,}98637 - 10}{10 + 7{,}315423} + 1} = 6{,}5856; \quad m_5 = 8{,}69087.$$

Далее вычисляем значения $c_\imath$:

$$c_1 = \frac{\Gamma(n_1 + m_1)}{(b-a)^{n_1+m_1-1} \cdot \Gamma(n_1) \cdot \Gamma(m_1)} =$$

$$= \frac{\Gamma(4{,}643588 + 10{,}633812)}{(22{,}98637 + 7{,}315423)^{4{,}643588+10{,}633812-1} \cdot \Gamma(6{,}643588) \cdot \Gamma(10{,}633812)} =$$

$$= \frac{\Gamma(15{,}2773)}{30{,}30179^{14{,}2773} \cdot \Gamma(4{,}643588) \cdot \Gamma(10б633812)} = \frac{1{,}8380547 \cdot 10^{11}}{30{,}30179^{14{,}2773} \cdot \Gamma(4{,}643588) \cdot \Gamma(10{,}633812)} =$$

$$= \frac{1{,}2971 \cdot 10^{-10}}{14{,}6994 \cdot 1541115{,}637} = 5{,}726 \cdot 10^{-18};$$

$$c_2 = \frac{1{,}2971 \cdot 10^{-10}}{\Gamma(5{,}0817) \cdot \Gamma(10{,}1956)} = \frac{1{,}2971 \cdot 10^{-10}}{27{,}1584 \cdot 565781{,}67} = 8{,}44 \cdot 10^{-18};$$

$$c_3 = \frac{1{,}2971 \cdot 10^{-10}}{\Gamma(5{,}9581) \cdot \Gamma(9{,}31929)} = \frac{1{,}2971 \cdot 10^{-10}}{111{,}71426 \cdot 80451{,}32} = 1{,}4431 \cdot 10^{-17};$$

$$c_4 = \frac{1{,}2971 \cdot 10^{-10}}{\Gamma(7{,}2726) \cdot \Gamma(8{,}0048)} = \frac{1.2971 \cdot 10^{-10}}{1208{,}4768 \cdot 5085{,}2762} = 2{,}1105 \cdot 10^{-17};$$

$$c_5 = \frac{1{,}2971 \cdot 10^{-10}}{\Gamma(6{,}5865) \cdot \Gamma(8{,}69087)} = \frac{1{,}2971 \cdot 10^{-10}}{335{,}9016 \cdot 20885{,}6629} = 1{,}8487 \cdot 10^{-17}.$$

Вычисляем значение функции вклада

$$\varphi_1(x_1) = c_1 \cdot (x_1 - a)^{n_1-1} \cdot (b - x_1)^{m_1-1} = \\ = 5{,}725 \cdot 10^{-18} \cdot (1 + 7{,}315423)^{3{,}643588} \cdot (22{,}98637 - 1)^{9{,}633812} = 0{,}109543.$$

Далее вычисляем функции вкладов по аналогии, используя различные значения x_j, c_i, n_i и m_i. Результаты вычислений сведены в таблицу:

$\varphi_i(x_j)$	$x_1 = 1$	$x_2 = 2$	$x_3 = 4$	$x_4 = 7$	$x_5 = 10$
$\varphi_1(x_j)$	0,109543	0,105809	0,081900	0,036800	0,009938
$\varphi_2(x_j)$	0,034305	0,009948	0,000064	$8 \cdot 10^{-9}$	$8{,}24 \cdot 10^{-6}$
$\varphi_3(x_j)$	0,019857	0,005677	0,000238	$4{,}887 \cdot 10^{-10}$	$3{,}541 \cdot 10^{-6}$
$\varphi_4(x_j)$	0,005578	0,001365	$3{,}5 \cdot 10^{-5}$	$4{,}103 \cdot 10^{-12}$	$6{,}52 \cdot 10^{-7}$
$\varphi_5(x_j)$	0,011499	0,003068	$1{,}01 \cdot 10^{-4}$	$5{,}31 \cdot 10^{-11}$	$1{,}68 \cdot 10^{-6}$
$\sum \varphi_i(x_j)$	0,180782	0,125867	0,082338	0,036800	0,009952
$f(x_j)$	0,036156	0,025173	0,016467	0,007360	0,001904

Там же приведены оценки плотности распределения вероятностей

$$f(x_j) = \frac{1}{n} \cdot \sum_{i=1}^{n} \varphi_i(x_j).$$

ГЛАВА 4

ПРОВЕРКА ГИПОТЕЗ О ЗНАЧЕНИЯХ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ

Общие положения. Высказывание гипотез о свойствах окружающего нас мира и проверка их непосредственными наблюдениями или с помощью целенаправленного экспериментирования составляет основу того, что́ мы называем наукой или научной деятельностью.

Оставляя в стороне проблему формирования гипотез (она лежит за пределами прикладной математической статистики), рассмотрим процесс проверки их справедливости по результатам наблюдений.

Проверка научной гипотезы применительно к потребностям ежедневной практики инженера и исследователя зависит от специфики наблюдаемых процессов и потребностей практики.

Например, часто встречающаяся на пути инженера задача — проверка соответствия параметров разработанного изделия предъявляемым требованиям — в математико-статистической формулировке может звучать так: „*необходимо проверить гипотезу о том, что параметр ε распределения случайной величины x превосходит заданную величину ε_0*".

Сколь разнообразен и сложен окружающий нас мир, столь многочисленны и разнообразны возможные гипотезы о его свойствах. Поэтому настоящая глава является наиболее объемным разделом настоящей книги. Она содержит в себе большое количество способов проверки гипотез и примеров их реализации.

Следует всегда помнить о противоречии между категоричностью и надежностью высказываний по гипотезе: как правило, надежное высказывание некатегорично, категорическое высказывание ненадежно. Мы выдвигаем гипотезу и отвергаем ее тогда, когда по выборке получаем результат, маловероятный при истинности выдвинутой гипотезы. Принятая граница маловероятности называется *уровнем значимости*. При превышении этого уровня выдвинутая нулевая гипотеза отклоняется. Иногда пользуются *доверительной вероятностью*, являющейся дополнением уровня значимости до 1. Различие между наблюдаемым и истинным значениями параметра распределения признается *значимым* на уровне значимости β, если правильная гипотеза будет отклонена в $\beta \cdot 100\%$ случаев (мы чаще пользуемся доверительной вероятностью $\alpha = 1 - \beta$).

При проверке гипотез возможны ошибки двух типов: первого рода — отклонение верной гипотезы, и второго рода — принятие ложной гипотезы.

Следует помнить, что уровень значимости (или доверительная вероятность) должен устанавливаться *перед*(!) получением данных.

Обычно на практике применяются доверительные вероятности 0,95 или 0,99 (уровни значимости 0,05 или 0,01). Важно неукоснительно выполнять основное требование — гипотезы должны быть выдвинуты *перед* статистическим анализом, сам *числовой материал не должен быть использован для выдвижения гипотезы.* Гипотезы, выдвинутые на основе анализа полученного материала, могут быть полезны только в качестве новых гипотез для последующих проверок.

4.1. Сравнение параметров распределений

Рассмотрим в качестве примера однопараметрическое распределение $f(x;\varepsilon)$ случайной величины x с параметром ε. Относительно ε можно выдвинуть три гипотезы (называемые нулевыми), в соответствии с которыми неизвестный параметр $\varepsilon = \varepsilon_0$, $\varepsilon > \varepsilon_0$ или $\varepsilon < \varepsilon_0$ (ε_0 — гипотетическое значение параметра). Символически эти гипотезы обозначаются как $H_0\colon \varepsilon = \varepsilon_0$; $H_0'\colon \varepsilon > \varepsilon_0$ и $H_0''\colon \varepsilon < \varepsilon_0$. Им противостоят альтернативы, соответственно записываемые по аналогии символами

$$H_1\colon \varepsilon \neq \varepsilon_0; \quad H_1'\colon \varepsilon \leqslant \varepsilon_0 \quad \text{и} \quad H_1''\colon \varepsilon \geqslant \varepsilon_0.$$

Основой для вынесения суждения по гипотезе является выборочная точечная оценка параметра, по которой строится доверительный интервал, включающий с заданной вероятностью неизвестное истинное значение параметра. Методы оценки параметров распределений подробно изложены в главе 2.

Между нахождением интервальных оценок параметров и проверкой гипотез об их возможных значениях существует тесная взаимосвязь. Это по сути два различных способа формулировки одной задачи. Для проверки гипотез о значениях параметра достаточно найти доверительный интервал параметра и проверить, попадает ли в него гипотетическое значение параметра.

Например, гипотеза $H_0\colon \varepsilon = \varepsilon_0$ не отклоняется с вероятностью α, если ε_0 попадает в двусторонний доверительный интервал параметра ε с коэффициентом доверия α. Гипотеза $H_0'\colon \varepsilon > \varepsilon_0$ не отклоняется, если ε_0 не превышает нижней границы одностороннего доверительного интервала для ε; гипотеза $H_0''\colon \varepsilon < \varepsilon_0$ не отклоняется, если ε_0 превосходит верхнюю границу одностороннего доверительного интервала.

Упрощенное изложение методологии проверки гипотез, изложенное выше, хотя и отражает суть проблемы, но, конечно же, не исчерпывает всех возможных ситуаций (например, когда нужно сравнить несколько (> 2) параметров одновременно и т. п.).

Следует помнить, что суждение по гипотезе приходится выносить по результатам выборочной реализации наблюдаемых случайных величин. Случайность наблюдаемых величин порождает, естественно, и некоторую неопределенность наших заключений по гипотезе. Привычная классическая формулировке „*результаты наблюдений подтверждают выдвинутую гипотезу*“ *должна уступить место формуле* „*результаты наблюдений с достоверностью α* (*заранее принятая величина*) *не противоречат выдвинутой гипотезе*“.

4.1.1. Сравнение параметров нормальных распределений

4.1.1.1. Сравнение двух средних значений

Имеются две выборки независимых случайных величин

$$x_1, x_2, \ldots, x_{n-1}, x_n \quad \text{и} \quad y_1, y_2, \ldots, y_{m-1}, y_m.$$

Необходимо на основе выборочных данных установить наличие значимой разницы в средних двух совокупностей, из которых извлечены выборки, т. е. проверить нулевую гипотезу $H_0\colon \mu_1 = \mu_2$ против альтернатив $H_1\colon \mu_1 \neq \mu_2$, $H_1'\colon \mu_1 > \mu_2$ и $H_1''\colon \mu_1 < \mu_2$.

4.1.1.1.1. Сравнение при известных дисперсиях σ_1^2 и σ_2^2

Статистика критерия проверки нулевой гипотезы имеет вид

$$z = (\bar{x} - \bar{y}) \left\{ \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right\}^{-\frac{1}{2}}, \quad \text{где} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad \bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i.$$

При справедливости нулевой гипотезы z-статистика имеет нормальное распределение. Гипотеза $H_0: \mu_1 = \mu_2$ предпочитается альтернативе $H_1: \mu_1 \neq \mu_2$ с доверительной вероятностью α, если $|z| < u_{\frac{1+\alpha}{2}}$; альтернативе $H_1': \mu_1 > \mu_2$, если $z < u_\alpha$; и альтернативе $H_1'': \mu_1 < \mu_2$, если $z > u_{1-\alpha}$ (здесь u_γ — γ-квантиль стандартного нормального распределения).

Задача 182. *Имеются две выборки случайных величин:*

$(n = 10)$ x_i: 1,2; 2,1; 3,2; 3,6; 3,8; 4,4; 6,1; 7,1; 9; 10,2;
$(m = 8)$ y_i: 2,4; 2,8; 4,1; 4,4; 6,8; 7,2; 8,9

с известными дисперсиями $\sigma_x^2 = 8{,}7$ и $\sigma_y^2 = 6{,}1$. Проверить гипотезу равенства средних значений $H_0: \mu_x = \mu_y$ в двух выборках при доверительной вероятности $\alpha = 0{,}95$ против альтернативы $H_1: \mu_1 \neq \mu_2$.

Вычисляем оценки $\bar{x} = \frac{1}{n} \cdot \sum x_i = 5{,}07$; $\bar{y} = \frac{1}{m} \cdot \sum y_i = 5{,}587$ и статистику проверки нулевой гипотезы $z = (5{,}07 - 5{,}587) \cdot \left\{ \frac{8{,}7}{10} + \frac{6{,}1}{8} \right\}^{-\frac{1}{2}} = -0{,}4046$.

Для $\alpha = 0{,}95$ имеем (см. табл. 1) значение $u_{\frac{1+\alpha}{2}} = u_{0{,}975} = 1{,}96$.

Так как $|z| = 0{,}4046 < u_{0{,}975} = 1{,}96$, нулевая гипотеза равенства средних не отклоняется.

4.1.1.1.2. Сравнение при неизвестных равных дисперсиях

Статистика критерия определяется формулой

$$t = \left(\frac{\bar{x} - \bar{y}}{s} \right) \left\{ \frac{m - n}{m + n} \right\}^{\frac{1}{2}},$$

где

$$s^2 = \frac{(n-1)\, s_1^2 + (m-1)\, s_2^2}{n + m - 2}; \quad s_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2; \quad s_2^2 = \frac{1}{m-1} \sum_{i=1}^{m} (y_i - \bar{y})^2.$$

При справедливости нулевой гипотезы статистика критерия имеет распределение Стьюдента (см. раздел 1.1.9) с $f = n + m - 2$ степенями свободы. Проверка нулевой гипотезы выполняется по аналогии со случаем, когда дисперсии известны, с заменой квантилей стандартного нормального распределения u_γ на квантили распределения Стьюдента t_γ, значения которых для $f \leqslant 30$ приведены в табл. 118 (при $f > 30$ можно принять $t_\gamma = u_\gamma$).

Задача 183. *Имеются два ряда выборочных данных:*

$(n = 12)$ x_i: 1, 2, 3, 5, 7, 12, 14, 16, 16, 17, 19, 22;
$(m = 10)$ y_i: 12, 16, 19, 22, 24, 26, 32, 34, 36, 44.

Необходимо проверить гипотезу равенства средних в обеих выборках $H_0: \mu_1 = \mu_2$ против альтернативы $H_1: \mu_1 \neq \mu_2$ при доверительной вероятности $\alpha = 0{,}90$.

Имеем $\bar{x} = 11{,}167$; $\bar{y} = 26{,}5$; $s_1^2 = 52{,}515$ $(s_1 = 7{,}247)$; $s_2^2 = 98{,}5$ $(s_2 = 9{,}925)$.

Проверкой по критерию Фишера (см. раздел 3.1.1.3.1) убеждаемся в статистической неразличимости (равенстве) значений s_1^2 и s_2^2. Вычисляем

$$s^2 = \frac{(12-1) \cdot 52{,}515 + (10-1) \cdot 98{,}5}{12 + 10 - 2} = 73{,}21 \ (s = 8{,}56);$$

$$z = \left(\frac{11{,}167 - 26{,}5}{8{,}56} \right) \cdot \left\{ \frac{12 + 10}{12 \cdot 10} \right\}^{-\frac{1}{2}} = -4{,}18.$$

Из табл. 118 для числа степеней свободы $f = 12 + 10 - 2 = 20$ и $\alpha = 0{,}90$ находим $t_{\frac{1+\alpha}{2}} = t_{0{,}95} = 1{,}725$. Так как $|t| = 4{,}18 > t_{0{,}95} = 1{,}725$, нулевая гипотеза равенства средних должна быть отклонена.

Таблица 118

Таблица квантилей распределения Стьюдента t_γ $(t_{1-\gamma} = -t_\gamma)$

f	γ				f	γ			
	0,90	0,95	0,975	0,99		0,90	0,95	0,975	0,99
1	3,078	6,314	12,706	31,821	16	1,337	1,746	2,120	2,583
2	1,886	2,920	4,303	6,695	17	1,333	1,740	2,110	2,567
3	1,638	2,353	3,182	4,541	18	1,330	1,734	2,101	2,552
4	1,533	2,132	2,776	3,747	19	1,328	1,729	2,093	2,539
5	1,476	2,015	2,571	3,365	20	1,325	1,725	2,086	2,528
6	1,440	1,943	2,447	3,143	21	1,323	1,721	2,080	2,518
7	1,415	1,895	2,365	2,998	22	1,321	1,717	2,074	2,508
8	1,397	1,860	2,306	2,896	23	1,319	1,714	2,069	2,500
9	1,383	1,833	2,262	2,821	24	1,318	1,711	2,064	2,492
10	1,372	1,812	2,228	2,764	25	1,316	1,708	2,059	2,485
11	1,363	1,796	2,201	2,718	26	1,315	1,706	2,056	2,479
12	1,356	1,782	2,179	2,681	27	1,314	1,703	2,052	2,473
13	1,350	1,771	2,160	2,650	28	1,313	1,701	2,048	2,467
14	1,345	1,761	2,145	2,624	29	1,311	1,699	2,045	2,462
15	1,341	1,753	2,131	2,602	30	1,310	1,697	2,042	2,457

4.1.1.1.3. Сравнение при неизвестных неравных дисперсиях

Задача сравнения средних двух нормально распределенных совокупностей при неизвестных и неравных (по выборочным оценкам) дисперсиях известна как проблема Беренса–Фишера [5] по имени авторов, впервые ее сформулировавших.

Точного решения этой задачи до настоящего времени нет. На практике обычно используются различные приближения, некоторые из которых рассмотрены ниже.

4.1.1.1.3.1. Критерий Кохрана–Кокса [386]

Статистика критерия $t_K = \dfrac{1}{s}(\bar{x} - \bar{y})$, где $s^2 = \dfrac{s_1^2}{n} + \dfrac{s_2^2}{m}$.

Критические значения статистики вычисляются по формуле

$$t'_\alpha = \frac{\nu_1 t_\alpha(f_1) + \nu_2 t_\alpha(f_2)}{\nu_1 + \nu_2}, \quad \text{где} \quad \nu_1 = \frac{s_1^2}{n} \quad \text{и} \quad \nu_2 = \frac{s_2^2}{m};$$

$t_\alpha(f)$ — α-квантиль распределения Стьюдента с f степенями свободы (в нашем случае $f_1 = n - 1$ и $f_2 = m - 1$).

4.1.1.1.3.2. Критерий Сатервайта [387]

Статистика критерия совпадает со статистикой критерия Кохрана—Кокс. Критическими значениями статистики являются квантили распределения Стьюдента с числом степеней свободы

$$f_C = s^4 \left\{ \frac{1}{n-1}\left(\frac{s_1^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{s_2^2}{m}\right)^2 \right\}^{-1}.$$

4.1.1.1.3.3. Критерий Уэлча [388]

Отличается от критерия Сатервайта только числом степеней свободы, при котором определяется критическое значение распределения Стьюдента. В критерии Уэлча

$$f_У = s^4 \left\{ \frac{1}{n+1} \left(\frac{s_1^2}{n} \right)^2 + \frac{1}{m-1} \left(\frac{s_2^2}{m} \right)^2 \right\}^{-1} - 2.$$

Задача 184. *Имеются две выборки данных:*

$(n = 10)$ x_i: 2, 4, 6, 7, 9, 12, 14, 16, 19, 24;
$(m = 9)$ y_i: 9, 14, 19, 21, 25, 29, 35, 41, 46.

Необходимо проверить гипотезу равенства средних при достоверности $\alpha = 0{,}95$.

Вычисляем $\bar{x} = 1{,}3$; $\bar{y} = 26{,}55$. Далее находим:

$$s_1^2 = \frac{1}{9} \cdot \sum_{i=1}^{10} (x_i - \bar{x})^2 = 49{,}12; \quad s_2^2 = \frac{1}{8} \cdot \sum_{i=1}^{9} (y_i - \bar{y})^2 = 152{,}528;$$

$$\nu_1 = \frac{s_1^2}{n} = 4{,}912; \quad \nu_2 = \frac{s_2^2}{m} = 16{,}947; \quad s^2 = \nu_1 + \nu_2 = 21{,}8595 \quad (s = 4{,}675).$$

Вычислим теперь статистики критериев для проверки нулевой гипотезы.

Критерий Кохрана–Кокс (4.1.1.1.3.1)

Находим по табл. 118 значения $t_{0,95}(f_1 = n - 1 = 9) = 1{,}833$ и $t_{0,95}(f_2 = m - 1 = 9 - 1 = 8) = 1{,}77$. Вычисляем статистику критерия и ее критическое значение:

$$t_К = \frac{26{,}55 - 11{,}33}{4{,}675} = 3{,}255; \quad t_{0,95} = \frac{4{,}912 \cdot 1{,}833 + 16{,}947 \cdot 1{,}86}{4{,}912 + 16{,}947} = 1{,}853.$$

Так как $t_К = 3{,}255 > t_{0,95} = 1{,}853$, нулевая гипотеза равенства средних отклоняется.

Критерий Сатервайта (4.1.1.1.3.2)

Вычисляем

$$t_С = 4{,}675 \cdot \left\{ \frac{1}{9} \cdot \left(\frac{4{,}912}{10} \right)^2 + \frac{1}{8} \cdot \left(\frac{152{,}528}{9} \right)^2 \right\}^{-1} = 12{,}57.$$

Из табл. 118 находим критическое значение: $t_{0,95}(f = 12{,}34) \approx 1{,}77$, которым также отклоняется нулевая гипотеза.

4.1.1.1.4. Модифицированный критерий Стьюдента [389]

Статистика критерия

$$T = 2 \frac{\bar{x} - \bar{y}}{\omega_1 + \omega_2},$$

где $\omega_1 = x_{\max} - x_{\min}$ и $\omega_2 = y_{\max} - y_{\min}$ — размахи сравниваемых выборок.

Критические значения T_α статистики для выборок одинакового объема $n = m$ приведены в табл. 119.

Если $|T| < T_{\frac{1+\alpha}{2}}$, то гипотеза $H_0\colon \mu_1 = \mu_2$ предпочитается альтернативе $H_1\colon \mu_1 \neq \mu_2$; если $T < T_\alpha$ — альтернативе $H_1'\colon \mu_1 > \mu_2$; если $T > T_\alpha$ — альтернативе $H_1''\colon \mu_1 < \mu_2$ (α — доверительная вероятность).

При $n \leqslant 10$ критерий не уступает по эффективности обычному критерию Стьюдента [390]. В связи с существенной потерей эффективности пользоваться критерием при $n \geqslant 20$ не рекомендуется.

Таблица 119

Критические значения модифицированного критерия Стьюдента для сравнения средних по двум выборкам равного объема [24]

n	Доверительная вероятность α				n	Доверительная вероятность α			
	0,95	0,975	0,99	0,995		0,95	0,975	0,99	0,995
2	2,322	3,427	5,553	7,916	12	0,214	0,260	0,315	0,355
3	0,974	1,272	1,715	2,093	13	0,201	0,243	0,294	0,331
4	0,644	0,813	1,047	1,237	14	0,189	0,228	0,276	0,311
5	0,493	0,613	0,772	0,896	15	0,179	0,216	0,261	0,293
6	0,405	0,499	0,621	0,714	16	0,170	0,205	0,247	0,278
7	0,347	0,426	0,525	0,600	17	0,162	0,195	0,236	0,264
8	0,306	0,373	0,459	0,521	18	0,155	0,187	0,225	0,252
9	0,275	0,334	0,409	0,464	19	0,149	0,179	0,216	0,242
10	0,250	0,304	0,371	0,419	20	0,143	0,172	0,207	0,232

Задача 185. *Имеются две выборки данных объема* $n = m = 10$:

$$x_{i1}:\ 2,\ 4,\ 6,\ 7,\ 9,\ 12,\ 14,\ 16,\ 19,\ 24;$$
$$x_{i2}:\ 9,\ 14,\ 19,\ 21,\ 25,\ 29,\ 35,\ 41,\ 46,\ 50.$$

Необходимо проверить гипотезу равенства средних модифицированным критерием Стьюдента при $\alpha = 0{,}95$ *против альтернативы* $H_1: \mu_1 \neq \mu_2$.

Имеем $\bar{x}_1 = 12{,}3$; $\bar{x}_2 = 28{,}9$; $\omega_1 = 24 - 2 = 22$; $\omega_2 = 50 - 9 = 41$.

Находим $T = 2 \cdot \dfrac{12{,}3 - 28{,}9}{22 + 41} = -0{,}527$.

Из табл. 119 для $\dfrac{1+\alpha}{2} = \dfrac{1+0{,}95}{2} = 0{,}975$ находим критическое значение T-статистики для $n = 10 - T_{0,975} = 0{,}304$.

Так как $|T| = 0{,}527 > T_{0,975} = 0{,}304$, нулевая гипотеза равенства средних отклоняется.

4.1.1.1.5. Парный t-критерий сравнения средних

Предположим, имеются две выборки случайных величин одинакового объема n, члены которых расположены в порядке их наблюдения. Требуется проверить гипотезу равенства средних в этих выборках. В такой постановке двухвыборочная задача может быть сформулирована как одновыборочная, если в качестве случайной величины рассматривать разности $y_i = x_{i1} - x_{i2}$ наблюдаемых пар

$$(x_{11}, x_{12}),\ \ldots,\ (x_{n1}, x_{n2}).$$

Статистика критерия аналогична обычной статистике Стьюдента (см. раздел 4.1.1.1.3)

$$t = \frac{\bar{y}}{s_y}\sqrt{n}, \quad \text{где} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i; \quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

Проверка нулевой гипотезы полностью аналогична обычному критерию Стьюдента.

Задача 186. *Проверить гипотезу равенства средних двух выборок в условиях задачи 185 парным t-критерием.*

Имеем последовательность разностей

$$y_i = x_{i1} - x_{i2} = 7,\ 10,\ 13,\ 14,\ 16,\ 17,\ 21,\ 25,\ 27,\ 26.$$

Находим

$$\bar{y} = 17{,}6; \quad s_y^2 = 48{,}044 \ (s_y = 6{,}931); \quad t = \frac{17{,}6}{6{,}931} \cdot \sqrt{10} = 8{,}03.$$

Из табл. 118 для $n = 10$ и $\alpha = 0{,}975$ находим $t_{0,975}(10) = 1{,}812$.
Так как $t = 8{,}03 > t_{0,975} = 1{,}812$, нулевая гипотеза отклоняется.

4.1.1.1.6. Критерий Уолша, основанный на порядковых статистиках [391]

Рассмотрим следующую ситуацию. Имеется выборка $x_1, \ldots, x_n$ объема n из нормального распределения со средним μ_1 и дисперсией σ^2. Получено одно независимое наблюдение y из нормального распределения со средним μ_2 и такой же дисперсией σ^2. Необходимо проверить гипотезу $H_0\colon \mu_1 = \mu_2$ против альтернатив $H_1'\colon \mu_1 < \mu_2$ и $H_1''\colon \mu_1 > \mu_2$. По сути, нулевая гипотеза утверждает, что независимая случайная величина y принадлежит распределению, из которого извлечена выборка $x_1, \ldots, x_n$.

Напомним, что k-й порядковой статистикой называется k-е по величине значение в выборке упорядоченных по возрастанию случайных величин $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$. Если выборочные значения упорядочены по возрастанию, то порядковые статистики просто являются элементами такой выборки.

Проверка гипотезы $H_0\colon \mu_1 = \mu_2$ включает в себя проверку справедливости серии неравенств.

Если

$$y - \bar{x} > \sqrt{n+1}\,(\bar{x} - x_m) \quad \text{или} \quad y + \bar{x} > \sqrt{n+1}\,(\bar{x} + x_{n+1-m}),$$

где $m = 1, \ldots, 4$, то нулевая гипотеза отклоняется в пользу альтернативы $H_1\colon \mu_2 > \mu_1$.

Если

$$y - \bar{x} < \sqrt{n+1}(\bar{x} - x_{n+1-m}) \quad \text{или} \quad y + \bar{x} < \sqrt{n+1}(\bar{x} + x_m),$$

то нулевая гипотеза отклоняется в пользу альтернативы $H_1'\colon \mu_1 > \mu_2$. В любом ином случае нулевая гипотеза не отклоняется. Уровень значимости α, принимаемый для проверки гипотезы, определяется в зависимости от принятого значения m следующим образом:

$$\alpha = \begin{cases} (0{,}5)^n & \text{при } m = 1; \\ (n+1)(0{,}5)^n & \text{при } m = 2; \\ (n^2+n+2)(0{,}5)^{n+1} & \text{при } m = 3; \\ \dfrac{1}{3}(n^3+5n+6)(0{,}5)^n & \text{при } m = 4. \end{cases}$$

По сравнению с обычным критерием Стьюдента (см. раздел 4.1.1.1.3) относительная эффективность критерия Уолша равна от 0,95 при $n = 6$ до 0,70 при $n = 16$.

Задача 187. *Имеется выборка данных*

$$x_i\colon 7,\ 10,\ 13,\ 14,\ 16,\ 17,\ 21,\ 26,\ 26,\ 27.$$

Проверить критерием Уолша гипотезу о принадлежности независимо полученного выборочного значения $y = 31$ распределению, из которого извлечена выборка x_i.

Имеем $\bar{x} = 17{,}6$ при $m = 1$ и $x_{10+1-1} = x_{10} = 27$.
Вычисляем

$$\sqrt{n+1} \cdot (\bar{x} - x_1) = 35{,}156; \qquad \sqrt{n+1} \cdot (\bar{x} + x_{10}) = 147{,}92;$$
$$\sqrt{n+1} \cdot (\bar{x} - x_{10}) = -31{,}176; \qquad \sqrt{n+1} \cdot (\bar{x} + x_1) = 81{,}589.$$

Непосредственной проверкой убеждаемся, что

$$y - \bar{x} = 31 - 17{,}6 = 13{,}4 > 35{,}156; \qquad y + \bar{x} = 31 + 17{,}6 = 48{,}6 < 147{,}02;$$
$$y - \bar{x} = 31 - 17{,}6 = 13{,}4 > -31{,}176; \quad y + \bar{x} = 31 + 17{,}6 = 48{,}6 < 81{,}589.$$

Видим, что одно из неравенств, а именно $y + \bar{x} < \sqrt{n+1} \cdot (\bar{x} + x_m)$, удовлетворяется, что позволяет отклонить нулевую гипотезу в пользу альтернативы $H_1: \mu_1 > \mu_2$. То есть значение $y = 31$ взято из распределения со средним, бóльшим, чем среднее распределения случайных величин x_i. Такой вывод сделан на уровне значимости $\alpha = (0{,}5)^{10} = 0{,}00097$, т. е. значимым признается событие, вероятность которого не превышает 0,00097.

Это очень низкий уровень значимости (обычно на практике используются значения 0,05 или 0,1). Для $\alpha = 0{,}05$ имеем $m = 3$, $x_3 = 13$ и $x_{10+1-3} = x_8 = 25$.

Тогда

$$\sqrt{n+1} \cdot (\bar{x} - x_3) = 15{,}25; \qquad \sqrt{n+1} \cdot (\bar{x} + x_8) = 141{,}28;$$
$$\sqrt{n+1} \cdot (\bar{x} - x_8) = -24{,}54; \quad \sqrt{n+1} \cdot (\bar{x} + x_3) = 101{,}48.$$

Убеждаемся, что неравенство $y + \bar{x} < \sqrt{n+1} \cdot (\bar{x} + x_3)$ выполняется и в этом случае, что и на уровне значимости $\alpha = 0{,}05$ отклоняет нулевую гипотезу в пользу гипотезы $H_1': \mu_1 > \mu_2$.

4.1.1.1.7. Двухступенчатый двухвыборочный медианный критерий Волфа [392]

Довольно любопытный критерий сравнения средних двух совокупностей, наиболее эффективно реализуемый применительно к испытаниям на долговечность технических объектов.

Процедура проверки гипотезы равенства двух средних заключается в следующем. Фиксируется выборка из первой совокупности объема n: $x_1, \ldots, x_n$. Определяется ее медиана $\tilde{x} = x_{\frac{n+1}{2}}$ (при n нечетном) или $\tilde{x} = \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$ (при n четном). Затем отбираются члены из второй совокупности y до тех пор, пока не будет получено ровно r значений y-ов, превосходящих медиану.

Статистикой критерия является число N_n значений y-ов, которое необходимо, чтобы получить ровно r значений y-ов, превосходящих медиану.

Таблица 120

Критические значения $N_0(r, n)$ для доверительной вероятности $\alpha = 0{,}95$ [392]

n	r	N_0	n	r	N_0	n	r	N_0	n	r	N_0
3	6	6	5	8	9	7	12	14	11	9	10
	7	7		9	10		13	15		10	12
	8	8		10	11	9	5	5	13	5	5
	9	9		11	12		6	6		6	6
	10	10		12	13		7	8		7	8
	11	12		13	15		8	9		8	9
	12	13		14	16		9	10	15	5	5
	13	14		15	17		10	12		7	8
	14	15	7	6	6		11	13		8	9
	15	16		7	7		12	14		9	11
	16	17		8	9	11	5	5	17	5	5
	17	18		9	10		6	6		6	7
5	6	6		10	11		7	8		7	8
	7	7		11	13		8	9		8	9

Нулевая гипотеза $H_0\colon \mu_y = \mu_x$ отклоняется в пользу альтернативы $H_1\colon \mu_y > \mu_x$ тогда и только тогда, когда $N_n \leqslant N_0(r,n)$, где $N_0(r,n)$ — критические значения, приведенные в табл. 120.

Таблица 121

Значения r и N_0 при $n \geqslant 20$ для двухступенчатого двухвыборочного критерия сравнения средних при доверительной вероятности $\alpha = 0{,}95$ [392]

r	4	5	6	7	8	9	10	11	12	13	14	15
N_0	4	5	7	8	10	12	13	15	17	18	20	21

При $n \geqslant 20$ рекомендуемые значения r и N_0 при $\alpha = 0{,}95$ приведены в табл. 121.

Задача 188. *При испытаниях $n = 11$ приборов получены значения ресурса*

x_i: 790, 830, 920, 1010, 1080, 1210, 1250, 1300, 1400, 4404 *ч.*

Необходимо проверить, увеличилась ли долговечность приборов после их усовершенствования.

Для проверки гипотезы были проведены испытания усовершенствованной партии приборов. В результате испытаний пяти приборов из усовершенствованной партии были получены следующие значения:

y_i: 1090, 1100, 1120, 1090, 1200 ч.

Исходя из табл. 120, для выбранного значения $r = 5$ необходимо, чтобы все первые пять значений ресурса превосходили медиану, равную в нашем случае 1080 ч. Убеждаемся, что $N_n = N_0$, что позволяет отклонить нулевую гипотезу в пользу альтернативы, утверждающей значимое смещение ресурса усовершенствованных приборов в сторону увеличения.

4.1.1.1.8. F-критерий для сравнения двух средних с одинаковыми дисперсиями

Критерий эквивалентен t-критерию (см. раздел 4.1.1.1.2) и использует соотношение между распределением Стьюдента и F-распределением $F_{1,f} = t_f^2$, где $F_{1,f}$ — случайная величина, имеющая распределение Фишера с $f_1 = 1$ и $f_2 = f$ степенями свободы.

Достоинством использования такой эквивалентной формы критерия является экономичность в вычислениях, объем которых снижается на $\approx 30\%$. Если мы имеем выборки $x_1, \ldots, x_{n_1}$ и $y_1, \ldots, y_{n_2}$, то статистика критерия проверки гипотезы $H_0\colon \mu_1 = \mu_2$ имеет вид

$$F = \frac{(n_1+n_2-1)\left(n_2\sum x_i - n_1\sum y_i\right)^2}{(n_1+n_2)\left[n_1n_2\left(\sum x_i^2+\sum y_i^2\right) - n_2\left(\sum x_1\right)^2 - n_1\left(\sum y_1\right)^2\right]}.$$

Величина F имеет распределение Фишера (см. раздел 1.1.10) с $f_1 = 1$ и $f_2 = n_1 + n_2 - 1$ степенями свободы. При $n_1 = n_2 = n$ имеет место соотношение

$$F = \frac{(n-1)\left(\sum x_i - \sum y_i\right)^2}{n\left(\sum x_i^2+\sum y_i^2\right) - \left[\left(\sum x_i\right)^2 + \left(\sum y_i\right)^2\right]}, \quad \text{где} \quad f_1 = 1, \quad f_2 = 2n-2.$$

Нулевая гипотеза отклоняется, если $F > F_\alpha(f_1, f_2)$, где $F_\alpha(f_1, f_2)$ — критическое значение F-статистики на уровне значимости α при f_1 и f_2 степенях свободы. Для его нахождения можно воспользоваться таблицами или аппроксимациями, приведенными в главе 1 (рекомендуется аппроксимация 9).

Задача 189. *Проверить F-критерием гипотезу равенства средних в двух выборках из нормального распределения при $\alpha = 0{,}95$ ($n_1 = n_2 = 30$)*

x_i:	1	3	4	5	7	8	11	14	19	21	23	26	30	35	41
	44	49	56	57	58	59	60	63	65	70	71	73	82	84	90;
y_i:	12	15	19	21	30	40	50	51	52	52	54	55	60	61	62
	64	64	65	70	71	73	75	79	80	81	84	85	86	87	90.

Имеем $n_1 = n_2 = 30$, $f_1 = 1$, $f_2 = 2 \cdot 30 - 2 = 58$.

Находим $\sum x_i = 1229$, $\sum y_i = 1788$; $\sum x_i^2 = 72805$; $\sum y_i^2 = 121146$ и вычисляем статистику критерия

$$F = \frac{(30-1)\cdot(1229-1788)^2}{30\cdot(72805+121146)-\left(1229^2+1788^2\right)} = 8{,}155.$$

Для $\alpha = 0{,}95$ находим из аппроксимации для F-распределения (см. аппроксимацию 9 в разделе 1.1.10) коэффициенты $a = 1{,}4287$, $b = 0{,}95$ и $c = 0{,}681$. Тогда имеем

$$h = \frac{2f_1f_2}{f_1+f_2} = 1{,}9661; \quad g = \frac{f_2-f_1}{f_1-f_2} = 0{,}98276;$$

$$\lg F_{0,95}(1;58) = 1{,}4287\cdot(1{,}9661-0{,}95)^{-\frac{1}{2}} - 0{,}681\cdot 0{,}98276 = 0{,}7709;$$

$$F_{0,95}(1;58) = 10^{0,7709} = 5{,}900.$$

Так как $F = 8,155 > F_{0,95}(1;58) = 5{,}900$, гипотеза равенства средних отклоняется.

4.1.1.2. Сравнение нескольких ($k > 2$) средних

Имеются k выборок равного объема n из нормально распределенной совокупности $x_{11}, x_{12}, \ldots, x_{1n_1}$; $x_{21}, x_{22}, \ldots, x_{2n_2}$; …; $x_{k1}, x_{k2}, \ldots, x_{kn_k}$.

Проверке подлежит нулевая гипотеза о статистической неразличимости средних $H_0\colon \mu_1 = \mu_2 = \ldots = \mu_k$ против альтернативы

$$H_1\colon\ |\mu_{i+1} - \mu_0| > 0 \quad (i = 1, \ldots, k).$$

4.1.1.2.1. Модифицированный критерий Стьюдента

Статистика критерия имеет вид

$$Q = \frac{\max\limits_{1\leqslant j\leqslant k}\sum\limits_{i=1}^{n} x_{ij} - \min\limits_{1\leqslant j\leqslant k}\sum\limits_{i=1}^{n} x_{ij}}{\sum\limits_{j=1}^{k}(x_{j\max} - x_{j\min})},$$

где $x_{j\max}$, $x_{j\min}$ — наибольшее и наименьшее значения j-й выборки.

Если $Q > Q_\alpha(k, n)$, где $Q_\alpha(k, n)$ — критическое значение статистики, приведенное в табл. 122, то нулевая гипотеза отклоняется.

Критерий достаточно прост в применении, но, в связи с большой потерей эффективности с ростом объема выборки, его рекомендуется применять при $n \leqslant 10$.

Таблица 122

Критические значения $Q_\alpha(k, n)$ модифицированного критерия Стьюдента [119, 393]

k	n								
	2	3	4	5	6	7	8	9	10
	Доверительная вероятность $\alpha = 0{,}05$								
1	18,10	7,60	6,20	5,80	5,70	5,70	5,80	5,90	6,00
2	3,50	2,50	2,30	2,30	2,40	2,40	2,50	2,60	2,60
3	1,79	1,44	1,41	1,43	1,47	1,52	1,57	1,62	1,68
4	1,18	1,01	1,01	1,03	1,07	1,11	1,14	1,19	1,23
5	0,89	0,78	0,78	0,81	0,84	0,87	0,91	0,94	0,97
6	0,71	0,64	0,64	0,66	0,69	0,71	0,74	0,77	0,80
7	0,59	0,53	0,54	0,56	0,59	0,61	0,64	0,66	0,68
8	0,50	0,46	0,47	0,49	0,51	0,53	0,56	0,58	0,60
9	0,44	0,41	0,42	0,43	0,45	0,47	0,49	0,51	0,53
10	0,39	0,36	0,37	0,39	0,40	0,42	0,44	0,46	0,47
	Доверительная вероятность $\alpha = 0{,}01$								
1	90,00	17,50	11,20	9,30	8,60	8,20	8,00	8,00	8,00
2	8,30	4,00	3,40	3,20	3,10	3,10	3,10	3,20	3,20
3	3,40	2,10	1,91	1,87	1,87	1,90	1,93	1,98	2,00
4	2,00	1,42	1,33	1,32	1,33	1,36	1,39	1,43	1,47
5	1,43	1,07	1,01	1,01	1,03	1,06	1,09	1,12	1,15
6	1,09	0,86	0,82	0,82	0,84	0,86	0,89	0,92	0,95
7	0,89	0,71	0,69	0,70	0,72	0,74	0,76	0,78	0,80
8	0,74	0,61	0,60	0,60	0,62	0,64	0,66	0,68	0,70
9	0,64	0,53	0,52	0,53	0,55	0,56	0,58	0,60	0,62
10	0,56	0,47	0,47	0,47	0,49	0,50	0,52	0,54	0,55

Задача 190. *В результате испытаний пяти выборок приборов объемом $n = 8$ каждая, изготовленных разными заводами, получены следующие значения долговечности приборов (ч):*

x_{1i}:	11	14	18	21	30	32	40	45
x_{2i}:	15	19	21	22	26	38	41	52
x_{3i}:	8	11	14	19	31	32	44	58
x_{4i}:	4	5	9	18	24	31	45	61
x_{5i}:	24	26	32	48	54	62	66	70.

Требуется проверить гипотезу о статистической неразличимости средних значений долговечности в выборках на уровне значимости $\alpha = 0{,}05$ модифицированным критерием Стьюдента.

Вычисляем предварительно суммы

$$\sum x_{1i} = 211; \quad \sum x_{2i} = 234; \quad \sum x_{3i} = 217; \quad \sum x_{4i} = 197; \quad \sum x_{5i} = 382;$$

$$\sum_{j=1}^{5} (x_{j\max} - x_{j\min}) = 234.$$

Имеем

$$\max_{i \leqslant j \leqslant k} \sum_{i=1}^{8} x_{ij} = 382; \quad \min_{i \leqslant j \leqslant k} \sum_{i=1}^{8} x_{ij} = 197; \quad Q = \frac{382 - 197}{224} = 0{,}8259.$$

Из табл. 122 для $\alpha = 0{,}05$, $k = 5$ и $n = 8$ находим критическое значение $Q_{0,05}(5{,}8) = 0{,}91$. Так как $Q = 0{,}826 < Q_{0,05}(5{,}8) = 0{,}91$, нулевая гипотеза равенства средних не отклоняется, однако критерий находится вблизи критической зоны.

4.1.1.2.2. Критерий „стьюдентизированного" размаха

Статистика критерия

$$q = \frac{\sqrt{n}}{s_f}\left(\max_{1\leqslant j\leqslant k}\bar{x}_j - \min_{1\leqslant j\leqslant k}\bar{x}_j\right),$$

где $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ и s_f — независимая оценка стандартного отклонения случайных величин x_{ij}, полученная на отдельной выборке объема $n = f + 1$.

Нулевая гипотеза отклоняется, если $q > q_\alpha(n,f)$, где $q_\alpha(n,f)$ — критическое значение, приведенное в табл. 123.

Для применения критерия необходимо предварительно иметь оценку стандартного отклонения σ по отдельной выборке и располагать информацией о том, что дисперсии во всех выборках одинаковы (методы проверки этого предположения, т. е. методы сравнения дисперсий, изложены в разделе 4.1.1.4).

Задача 191. *Проверить на уровне значимости $\alpha = 0{,}95$ гипотезу равенства средних критерием „стьюдентизированного" размаха, если оценка стандартного отклонения s_f получена по выборке*

$$x_i\colon\ 10,\ 12,\ 14,\ 19,\ 22,\ 28,\ 34,\ 46,\ 52,\ 61,\ 70.$$

В качестве выборочного материала использовать данные задачи 190.

Имеем

$$\bar{x}_1 = 26{,}375;\quad \bar{x}_2 = 29{,}25;\quad \bar{x}_3 = 27{,}125;\quad \bar{x}_4 = 24{,}625;\quad \bar{x}_5 = 47{,}75;$$

$$\max\bar{x}_j = 47{,}75;\quad \min\bar{x}_j = 26{,}375;\quad s_f = \left\{\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{1}{2}} = 20{,}868.$$

Тогда $q = \dfrac{\sqrt{10}}{20{,}868}\cdot(47{,}75 - 26{,}375) = 3{,}239$.

Из табл. 123 для $\alpha = 0{,}05$, $f = 10$ и $n = 8$ имеем $q_{0{,}05}(8,10) = 5{,}30$.

Так как $q = 3{,}239 < q_{0{,}05}(8,10) = 5{,}30$, гипотеза равенства средних не отклоняется.

4.1.1.2.3. Дисперсионный критерий

Статистика критерия имеет вид

$$F = \frac{kn\,(n-1)}{k-1}\,\frac{\sum\limits_{j=1}^{k}(\bar{x}_j - \bar{x})^2}{\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{n}(x_{ij} - \bar{x}_j)^2},$$

где

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij};\qquad \bar{x} = \frac{1}{kn}\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}.$$

При справедливости нулевой гипотезы статистика критерия имеет распределение Фишера (см. раздел 1.1.10) с $f_1 = k - 1$ и $f_2 = k(n-1)$ степенями свободы.

При $F > F_\alpha(k-1; k\,(n-1))$ нулевая гипотеза отклоняется. Здесь $F_\alpha(f_1, f_2)$ — α-квантиль F-распределения, ее значения могут быть найдены или по соответствующим таблицам, или с помощью аппроксимации.

Таблица 123

Критические значения $q_\alpha(n, f)$ „стьюдентизированного“ размаха [119]

n	f								
	2	5	10	15	20	30	40	60	∞
Доверительная вероятность $\alpha = 0{,}05$									
2	6,08	3,64	3,15	3,01	2,95	2,89	2,86	2,83	2,77
3	8,33	4,60	3,88	3,67	3,58	3,49	3,44	3,40	3,31
4	9,80	5,22	4,33	4,08	3,96	3,85	3,79	3,74	3,63
5	10,88	5,67	4,65	4,37	4,23	4,10	4,04	3,98	3,86
6	11,74	6,03	4,91	4,59	4,45	4,30	4,23	4,16	4,03
7	12,44	6,33	5,12	4,78	4,62	4,46	4,39	4,31	4,17
8	13,03	6,58	5,30	4,94	4,77	4,60	4,52	4,44	4,29
9	13,54	6,80	5,46	5,08	4,90	4,72	4,63	4,55	4,39
10	13,99	6,99	5,60	5,20	5,01	4,82	4,73	4,65	4,47
11	14,39	7,17	5,72	5,31	5,11	4,92	4,82	4,73	4,55
12	14,75	7,32	5,83	5,40	5,20	5,00	4,90	4,81	4,62
13	15,08	7,47	5,93	5,49	5,28	5,08	4,98	4,88	4,68
14	15,38	7,60	6,03	5,57	5,36	5,15	5,04	4,94	4,73
15	15,65	7,72	6,11	5,65	5,43	5,21	5,11	5,00	4,80
16	15,91	7,83	6,19	5,72	5,49	5,27	5,16	5,06	4,85
17	16,14	7,93	6,27	5,78	5,55	5,33	5,22	5,11	4,89
18	16,37	8,03	6,34	5,85	5,61	5,38	5,27	5,15	4,93
19	16,57	8,12	6,40	5,90	5,66	5,43	5,31	5,20	4,97
20	16,77	8,21	6,47	5,96	5,71	5,47	5,36	5,24	5,01
Доверительная вероятность $\alpha = 0{,}01$									
2	14,04	5,70	4,48	4,17	4,02	3,89	3,82	3,76	3,64
3	19,02	6,98	5,27	4,84	4,64	4,45	4,37	4,28	4,12
4	22,29	7,80	5,77	5,25	5,02	4,80	4,70	4,59	4,40
5	24,72	8,42	6,14	5,56	5,29	5,05	4,93	4,82	4,60
6	26,63	8,91	6,43	5,80	5,51	5,24	5,11	4,99	4,76
7	28,20	9,32	6,67	5,99	5,69	5,40	5,26	5,13	4,88
8	29,53	9,67	6,87	6,16	5,84	5,54	5,39	5,25	4,99
9	30,68	9,97	7,05	6,31	5,97	5,65	5,50	5,36	5,08
10	31,69	10,24	7,21	6,44	6,09	5,76	5,60	5,45	5,16
11	32,59	10,48	7,36	6,55	6,19	5,85	5,69	5,53	5,23
12	33,40	10,70	7,49	6,66	6,28	5,93	5,76	5,60	5,29
13	34,13	10,89	7,60	6,76	6,37	6,01	5,83	5,67	5,35
14	34,81	11,08	7,71	6,84	6,45	6,08	5,90	5,73	5,40
15	35,43	11,24	7,81	6,93	6,52	6,14	5,96	5,78	5,45
16	36,00	11,40	7,91	7,00	6,59	6,20	6,02	5,84	5,49
17	36,53	11,55	7,99	7,07	6,65	6,26	6,07	5,89	5,54
18	37,03	11,68	8,08	7,14	6,71	6,31	6,12	5,93	5,57
19	37,50	11,81	8,15	7,20	6,77	6,36	6,16	5,97	5,61
20	37,95	11,93	8,23	7,26	6,82	6,41	6,21	6,01	5,65

При $f_2 = k(n-1) > 4$ можно использовать упрощенный критерий Романовского [13], основанный на статистике

$$R = \frac{|Q-1|}{\sigma_Q},$$

где

$$Q = \frac{n\,[k\,(n-1)-2]}{k-1}\,\frac{\sum\limits_{j=1}^{k}(\bar{x}_j-\bar{x})^2}{\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{n}(x_{ij}-\bar{x}_j)^2};\quad \sigma_Q = \sqrt{\frac{2\,(kn-3)}{(k-1)\,(kn-k-4)}}.$$

Если $R \geqslant 3$, нулевая гипотеза отклоняется.

Изложенная процедура называется однофакторным дисперсионным анализом. Подробно дисперсионный анализ изложен в главе 5.

Следует помнить, что применение этого критерия при отсутствии нормальности исходных распределений величин x_{ij} не рекомендуется, так как он становится в этом случае неустойчивым.

Его устойчивость к отклонениям от нормальности повышается, если использовать модифицированные степени свободы для F-критерия [87]:

$$f_1 = d(k-1) \quad \text{и} \quad f_2 = dk(n-1),$$

где

$$d = 1 + \frac{kn+1}{kn-1}\frac{c}{kn-c};\quad c = \frac{\lambda}{\nu^2};\quad \nu = \frac{s_2}{kn-1};$$

$$\lambda = \frac{kn(kn+1)s_4 - 3(kn-1)s_2^2}{(kn-1)(kn-2)(kn-3)};\quad s_m = \sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij}-\bar{x})^m.$$

Задача 192. *Проверить в условиях задачи* 190 *гипотезу равенства средних при доверительной вероятности* $\alpha = 0{,}95$.

Имеем

$$\bar{x}_1 = 26{,}375;\quad \bar{x}_2 = 29{,}25; \bar{x}_3 = 27{,}125;\quad \bar{x}_4 = 24{,}625;\quad \bar{x}_5 = 47{,}75;$$

$$\bar{x} = \frac{1}{5\cdot 8}\cdot\sum_{j=1}^{5}\sum_{i=1}^{8}x_{ij} = 31{,}025;\quad \sum_{j=1}^{5}(\bar{x}_j - x)^2 = 363{,}66875;\quad \sum_{j=1}^{5}\sum_{i=1}^{8}(x_{ij}-\bar{x}_j)^2 = 9603{,}124.$$

Отсюда $F = \dfrac{5\cdot 8\cdot 7}{5-1}\cdot\dfrac{363{,}66875}{9603{,}124} = 2{,}65$.

Для $\alpha = 0{,}95$, $f_1 = 5-1 = 4$ и $f_2 = 5\cdot(8-1) = 35$ имеем $F_{0{,}95}(4;35) = 2{,}67$ (см., например, табл. А7 в [10]).

Так как $F = 2{,}65 < F_{0{,}95}(4;35) = 2{,}67$, нулевая гипотеза не отклоняется.

Рассмотрим теперь более устойчивый критерий. Вычисляем

$$s_2 = \sum_{j=1}^{5}\sum_{i=1}^{8}(x_{ij}-\bar{x}_j)^2 = 9602{,}34;\quad s_4 = \sum_{j=1}^{5}\sum_{i=1}^{8}(x_{ij}-\bar{x}_j)^4 = 5077427{,}388;$$

$$\lambda = \frac{5\cdot 8\cdot(5\cdot 8+1)\cdot 5077427{,}388 - 3\cdot(5\cdot 8-1)\cdot 9602{,}34^2}{(5\cdot 8-1)\cdot(5\cdot 8-2)\cdot(5\cdot 8-3)} = -44880{,}84582;$$

$$\nu = \frac{9602{,}34}{5\cdot 8-1} = 246{,}2138; c = -\frac{44880{,}84582}{246{,}2138^2} = -0{,}7403;$$

$$d = 1 + \frac{5\cdot 8+1}{5\cdot 8-1}\cdot\frac{(-0{,}7403)}{5\cdot 8+0{,}7403} = 0{,}98089;$$

$$f_1 = 0{,}98089\cdot(5-1) = 3{,}923;\quad f_2 = 0{,}98089\cdot 5\cdot 7 = 34{,}3.$$

Видим, что число степеней свободы несколько снижается, что позволяет отклонить гипотезу.

Рассмотрим упрощенный критерий Романовского:

$$Q = \frac{8\cdot(5\cdot 7-2)}{5-1}\cdot\frac{360{,}66875}{9602{,}234} = 2{,}479; \quad \sigma_Q = \sqrt{\frac{2\cdot(5\cdot 8-3)}{(5-1)\cdot(5\cdot 8-5-4)}} = 0{,}7725;$$

$$R = \frac{2{,}479}{0{,}7725} = 3{,}209.$$

Так как $R > 3$, нулевая гипотеза отклоняется.

4.1.1.2.4. Критерий Полсона

В [394] Полсоном рассмотрена проблема выделения среди k выборок, по n наблюдений в каждой, выборки со средним значением, бóльшим, чем у $(k-1)$ остальных.

Статистика критерия имеет вид

$$\lambda = \frac{n\left(\max\limits_{1\leqslant j\leqslant k}\bar{x}_j - \bar{x}\right)}{\sqrt{\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{n}(x_{ij}-\bar{x})^2}}, \quad \text{где} \quad \bar{x} = \frac{1}{k}\sum_{j=1}^{k}\bar{x}_j.$$

Если $\lambda \leqslant \lambda_\alpha$, то с вероятностью α справедлива нулевая гипотеза $H_o\colon\ \bar{x}_1 = \ldots = \bar{x}_k$. В ином случае $(\lambda > \lambda_\alpha)$ выборка с наибольшим средним признается значимо отличной от остальных.

Критическое значение статистики равно

$$\lambda_\alpha = \left\{\frac{n(k-1)F_{\alpha'}}{k[k(n-1)+k-2+F_{\alpha'}]}\right\}^{\frac{1}{2}},$$

где $F_{\alpha'}$ — α'-квантиль распределения Фишера с $f_1 = 1$ и $f_2 = kn-2$ степенями свободы и $\alpha' = \dfrac{2\alpha}{k}$.

Значения для $\alpha' = \dfrac{2\alpha}{k}$ следует брать из обычных таблиц F-распределения, используя аппроксимацию по α. Некоторые значения $F_{\alpha'}(1; kn-2)$ приведены в табл. 124, составленной на основе данных таблиц t-статистики Бонферрони [87, 395].

Задача 193. *Проверить критерием Полсона в условиях задачи* 190 *гипотезу равенства средних против альтернативы, утверждающей, что выборка с наибольшим средним значимо отличается от остальных* $(\alpha = 0{,}95)$.

Имеем $\max\limits_{1\leqslant j\leqslant k}\bar{x}_j = 47{,}75$; $\bar{x} = 31{,}025$; $\sum\limits_{j=1}^{5}\sum\limits_{i=1}^{8}(x_{ij}-\bar{x})^2 = 9602{,}34$.

Находим $\lambda = \dfrac{8\cdot(47{,}75-31{,}025)}{\sqrt{9602{,}34}} = 1{,}365$. Из табл. 124 для $k=5$, $n=8$ и $\alpha = 0{,}95$ имеем $F'_\alpha = 5{,}90$. Вычисляем критическое значение

$$\lambda_{0{,}95} = \sqrt{\frac{8\cdot(5-1)\cdot 5{,}90}{5\cdot(5\cdot 7+5-2+5{,}90)}} = 0{,}927.$$

Так как $\lambda = 1{,}365 > \lambda_{0{,}95} = 0{,}927$, нулевая гипотеза отклоняется и выборка со средним $\bar{x}_j = 47{,}75$ должна быть признана значимо отличающейся от остальных. Этот вывод не совпадает с ранее полученным, что может быть следствием отклонения распределения значений x_{ij} от нормального, а критерий Полсона очень критичен к нормальности распределения.

Таблица 124

Значения $F_{\alpha'}(1; kn-2)$ для доверительной вероятности $\alpha = 0{,}95$

n	k								
	2	3	4	5	6	7	8	9	10
3	7,71	7,10	6,94	7,02	7,61	7,31	7,54	7,29	7,63
4	5,99	5,91	6,30	6,51	7,11	6,93	7,16	7,30	7,35
5	5,32	5,68	5,98	6,25	6,86	6,72	7,05	7,15	7,24
6	4,96	5,41	5,79	6,09	6,70	6,59	6,91	7,11	7,05
7	4,75	5,29	5,66	5,98	6,59	6,49	6,85	6,95	6,95
8	4,60	5,17	5,57	5,90	6,54	6,42	6,81	6,90	6,91
9	4,49	5,09	5,50	5,86	6,49	6,39	6,79	6,86	6,87
10	4,41	5,01	5,45	5,81	6,45	6,37	6,77	6,82	6,85
11	4,35	4,96	5,41	5,78	6,39	6,34	6,75	6,78	6,83
12	4,30	4,93	5,39	5,75	6,37	6,30	6,72	6,75	6,81
13	4,26	4,89	5,35	5,71	6,35	6,28	6,70	6,72	6,79
14	4,22	4,85	5,34	5,68	6,33	6,26	6,68	6,69	6,77
15	4,20	4,83	5,29	5,66	6,31	6,24	6,65	6,67	6,75
16	4,17	4,81	5,26	5,63	6,29	6,22	6,62	6,65	6,73
17	4,15	4,79	5,24	5,61	6,27	6,20	6,59	6,63	6,71
18	4,13	4,77	5,21	5,59	6,25	6,18	6,57	6,60	6,70
19	4,11	4,76	5,18	5,57	6,23	6,16	6,55	6,59	6,68
20	4,10	4,75	5,15	5,56	6,20	6,15	6,53	6,58	6,63
30	4,00	4,63	5,02	5,41	6,02	6,02	6,41	6,51	6,52

4.1.1.2.5. Метод прямого сравнения (критерий Тьюки)

В [396] рассмотрен критерий, основанный на последовательности статистик

$$T_j = \frac{|\bar{x}_j - \bar{x}|}{s\sqrt{\dfrac{k-1}{kn}}},$$

сравнивающих попарно все исследуемые средние $\bar{x}_j$ с общим средним. В этом случае s^2 является оценкой общей дисперсии с $f = k(n-1)$ степенями свободы, т. е.

$$s^2 = \frac{1}{k\,(n-1)} \sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{x})^2 .$$

Если $T_j < T_\alpha$ для всех $j = 1, \ldots, k$, где T_α — критическое значение статистики, приведенное в табл. 125, то нулевая гипотеза не отклоняется. Нарушение неравенства при любом значении j отклоняет нулевую гипотезу. Предполагается, что дисперсии s_j^2 всех выборок статистически неразличимы. Этот критерий является альтернативой дисперсионному анализу (см. раздел 4.1.1.2.3).

Задача 194. *В условиях задачи* 190 *проверить гипотезу равенства средних методом прямого сравнения при доверительной вероятности* $\alpha = 0{,}95$.

Имеем

$$\bar{x}_1 = 26{,}375; \quad \bar{x}_2 = 29{,}25; \quad \bar{x}_3 = 27{,}125; \quad \bar{x}_4 = 24{,}625; \quad \bar{x}_5 = 31{,}025;$$

$$s_2 = \frac{1}{5 \cdot (8-1)} \cdot \sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{x})^2 = \frac{1}{35} \cdot 9602{,}34 = 274{,}35257.$$

Таблица 125

Критические значения T_α критерия Тьюки [396]

f	k							
	3	4	5	6	7	8	9	10
Доверительная вероятность $\alpha = 0{,}90$								
3	3,164							
4	2,813	3,094						
5	2,629	2,877	3,054					
6	2,516	2,744	3,032					
7	2,440	2,654	2,806	2,924	3,020			
8	2,385	2,589	2,734	2,846	2,937	3,015		
9	2,344	2,540	2,679	2,787	2,875	2,949	3,014	
10	2,312	2,502	2,636	2,740	2,826	2,898	2,960	3,015
11	2,286	2,472	2,602	2,703	2,786	2,856	2,917	2,970
12	2,265	2,447	2,574	2,673	2,754	2,822	2,881	2,934
13	2,247	2,426	2,551	2,647	2,727	2,794	2,852	2,903
14	2,232	2,408	2,531	2,626	2,704	2,769	2,826	2,876
15	2,219	2,393	2,514	2,607	2,684	2,748	2,804	2,854
16	2,208	2,379	2,499	2,591	2,667	2,730	2,786	2,834
17	2,198	2,368	2,486	2,577	2,652	2,715	2,769	2,817
18	2,190	2,357	2,474	2,565	2,638	2,700	2,754	2,802
19	2,182	2,348	2,464	2,553	2,626	2,688	2,741	2,788
20	2,175	2,340	2,455	2,544	2,616	2,677	2,730	2,776
30	2,133	2,290	2,398	2,482	2,550	2,607	2,657	2,700
60	2,092	2,241	2,343	2,422	2,486	2,539	2,586	2,627
120	2,068	2,193	2,289	2,363	2,423	2,473	2,516	2,554
Доверительная вероятность $\alpha = 0{,}95$								
3	4,148							
4	3,565	3,889						
5	3,254	3,527	3,724					
6	3,068	3,311	3,486	3,622				
7	2,944	3,168	3,327	4,453	3,556			
8	2,857	3,066	3,215	3,332	3,428	3,510		
9	2,791	2,990	3,131	3,242	3,333	3,410	3,478	
10	2,741	2,931	3,066	3,172	3,259	3,333	3,397	3,454
11	2,700	2,884	3,014	3,116	3,200	3,271	3,333	3,388
12	2,667	2,845	2,971	3,070	3,151	3,220	3,281	3,334
13	2,640	2,813	2,936	3,032	3,111	3,178	3,237	3,289
14	2,617	2,786	2,906	3,000	3,077	3,143	3,200	3,250
15	2,597	2,763	2,881	2,973	3,048	3,112	3,168	3,219
16	2,580	2,743	2,859	2,949	3,023	3,086	3,141	3,189
17	2,565	2,726	2,840	2,928	3,001	3,063	3,117	3,164
18	2,552	2,711	2,823	2,910	2,982	3,043	3,096	3,143
19	2,540	2,697	2,808	2,894	2,965	3,025	3,077	3,123
20	2,530	2,685	2,794	2,879	2,949	3,009	3,060	3,106
30	2,465	2,610	2,711	2,790	2,854	2,909	2,957	2,998
60	2,404	2,538	2,632	2,704	2,764	2,814	2,857	2,896
120	2,367	2,497	2,587	2,656	2,713	2,761	2,803	2,839

Вычисляем

$$T_1 = \frac{|26{,}375 - 31{,}025|}{16{,}562 \cdot \dfrac{5-1}{5 \cdot 8}} = 0{,}888; \quad T_2 = \frac{|29{,}25 - 31{,}025|}{5{,}23768} = 0{,}339;$$

$$T_3 = \frac{|27{,}125 - 31{,}025|}{5{,}23768} = 0{,}745; \quad T_4 = \frac{|24{,}625 - 31{,}025|}{5{,}23768} = 1{,}222;$$

$$T_5 = \frac{|47{,}75 - 31{,}025|}{5{,}23768} = 3{,}193.$$

Для $f = k \cdot (n-1) = 5 \cdot (8-1) = 35$ степеней свободы, $\alpha = 0{,}95$ и $k = 5$ из табл. 125 имеем $T_{0,95} = 2{,}6$. Так как $T_5 = 3{,}193 > T_{0,95} = 2{,}6$, нулевая гипотеза равенства средних отклоняется, так как $\bar{x}_5$ значимо отклоняется от $\bar{x}$.

Этот вывод совпадает с заключением, полученным при использовании критерия Полсона (задача 193).

4.1.1.2.6. Критерий „стьюдентизированного“ максимума (обобщенный критерий Тьюки)

Критерий основан на использовании верхней $\alpha \cdot 100\%$-й критической точки $m_{\alpha,k^*,\nu}$ модуля „стьюдентизированного“ максимума $\max\left|\dfrac{x_i}{s}\right|$ с параметрами $k^* = \dfrac{k\,(k-1)}{2}$ и $\nu = k\,(n-1)$. Подробно критерий рассмотрен в [397, 398].

Средние i-й и j-й выборок признаются не различающимися, если

$$M_{ij} = \frac{|\bar{x}_i - \bar{x}_j|}{s\sqrt{\dfrac{n}{2}}} < m_{\alpha,k^*,\nu} \quad (i, j = 1, \ldots, k;\ i \neq j).$$

Здесь, как и ранее, s^2 является оценкой дисперсии объединенной выборки с $\nu = k(n-1)$ степенями свободы, т. е.

$$s^2 = \frac{1}{k\,(n-1)} \sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{x})^2.$$

Таким образом, нулевая гипотеза равенства всех $j = 1, \ldots, k$ средних не отклоняется только тогда, когда все $\dfrac{k(k-1)}{2}$ пар средних удовлетворяют вышеприведенному неравенству.

Таблицы значений $m_{\alpha,k^*,\nu}$ опубликованы в [399, 400], наиболее полные таблицы содержатся в [398]. Фрагмент этой таблицы воспроизведен в табл. 126.

Задача 195. *В условиях задачи* 190 *проверить гипотезу равенства средних критерием „стьюдентизированного“ максимума при доверительной вероятности* $\alpha = 0{,}95$.

Имеем

$$n = 8; \quad k = 5; \quad \nu = k \cdot (n-1) = 35; \quad \bar{x}_1 = 26{,}375; \quad \bar{x}_2 = 29{,}25; \quad \bar{x}_3 = 27{,}125;$$
$$\bar{x}_4 = 24{,}625; \quad \bar{x}_5 = 47{,}75; s = 16{,}563.$$

Вычисляем $\dfrac{k \cdot (k-1)}{2} = \dfrac{5 \cdot 4}{3} = 10$ значений M_{ij}:

$$M_{12} = \frac{|\bar{x}_1 - \bar{x}_2|}{s \cdot \sqrt{\dfrac{2}{n}}} = \frac{|26{,}375 - 29{,}25|}{16{,}563 \cdot \sqrt{\dfrac{2}{8}}} = 0{,}347$$

и далее по аналогии

$$M_{13} = 0{,}091; \quad M_{14} = 0{,}212; \quad M_{15} = 2{,}585; \quad M_{23} = 0{,}257; \quad M_{24} = 0{,}559;$$
$$M_{25} = 2{,}237; \quad M_{34} = 0{,}302; \quad M_{35} = 2{,}494; \quad M_{45} = 2{,}797.$$

Из табл. 126 для $\nu = k \cdot (n-1) = 35$, $k = 5$ и $\alpha = 0{,}95$ получаем (интерполяцией) $m_{\alpha,k^*,\nu} = 2{,}97$. Так как все значения $M_{ij} < 2{,}97$, следует принять нулевую гипотезу равенства средних. Необходимо помнить, что применение этого критерия предполагает равенство (стохастическое) выборочных дисперсий сравниваемых выборок.

Таблица 126

Критические значения критерия $m_{\alpha,k^*,\nu}$ „стьюдентизированного“ максимума [398]

k	k^*	ν									
		5	7	10	12	16	20	24	30	60	120
Доверительная вероятность $\alpha = 0{,}95$											
3	3	3,40	3,05	2,83	2,75	2,65	2,59	2,56	0,52	2,45	2,42
4	6	3,93	3,49	3,20	3,09	2,97	2,90	2,85	2,80	2,76	2,72
5	10	4,31	3,80	3,47	3,34	3,20	3,11	3,06	3,00	2,90	2,85
6	15	4,61	4,05	3,68	3,54	3,38	3,28	3,22	3,16	3,04	2,98
7	21	4,85	4,25	3,85	3,77	3,52	3,42	3,35	3,29	3,16	3,09
8	28	5,06	4,42	3,99	3,83	3,64	3,53	3,46	3,39	3,25	3,18
9	36	5,23	4,57	4,12	3,95	3,75	3,63	3,56	3,48	3,33	3,19
10	45	5,38	4,69	4,22	4,05	3,84	3,72	3,64	3,56	3,41	3,33
12	66	5,64	4,91	4,41	4,22	4,00	3,87	3,78	3,70	3,53	3,44
14	91	5,85	5,08	4,56	4,37	4,13	3,99	3,90	3,81	3,63	3,54
16	120	6,03	5,23	4,69	4,49	4,24	4,09	4,00	3,90	3,71	3,62
18	153	6,18	5,36	4,80	4,59	4,33	4,18	4,08	3,98	3,78	3,68
20	190	6,31	5,47	4,90	4,68	4,42	4,26	4,16	4,06	3,85	3,75
Доверительная вероятность $\alpha = 0{,}90$											
3	3	2,78	2,55	2,41	2,36	2,29	2,25	2,23	2,21	2,16	2,14
4	6	3,24	2,96	2,77	2,70	2,62	2,57	2,53	2,50	2,44	2,41
5	10	3,58	3,25	3,03	2,95	2,84	2,79	2,75	2,71	2,63	2,60
6	15	3,84	3,48	3,23	3,14	3,02	2,96	2,91	2,87	2,78	2,74
7	21	4,05	3,66	3,39	3,29	3,16	3,09	3,04	3,00	2,90	2,85
8	28	4,22	3,81	3,53	3,42	3,29	3,21	3,16	3,10	3,00	2,95
9	36	4,37	3,94	3,64	3,53	3,39	3,31	3,25	3,20	3,09	3,03
10	45	4,51	4,06	3,75	3,63	3,48	3,39	3,33	3,28	3,16	3,10
12	66	4,73	4,25	3,92	3,79	3,63	3,54	3,48	3,41	3,29	3,22
14	91	4,91	4,41	4,06	3,92	3,76	3,66	3,59	3,52	3,39	3,32
16	120	5,06	4,54	4,18	4,04	3,86	3,76	3,69	3,62	3,47	3,40
18	153	5,19	4,66	4,28	4,13	3,95	3,85	3,77	3,70	3,55	3,47
20	190	5,30	4,76	4,37	4,22	4,03	3,92	3,85	3,77	3,62	3,54

4.1.1.2.7. Критерий Шеффе

Среди k средних значений, предварительно упорядоченных по величине: $\bar{x}_1 \leqslant \bar{x}_2 \leqslant \ldots \leqslant \bar{x}_k$, производится $(k-1)$ сравнений. Например, для $k = 5$ производится $k - 1 = 4$ множественных сравнения

$$\bar{x}_5 - \bar{x}_1, \quad \bar{x}_5 - \bar{x}_2, \quad \bar{x}_5 - \bar{x}_3, \quad \bar{x}_5 - \bar{x}_4.$$

Если при этом будет превышена критическая разница, то нулевая гипотеза о равенстве средних μ_i и μ_j отклоняется.

Шеффе [401] предложил использовать в качестве критического значения величину

$$D_1 = s\sqrt{\frac{2\,(k-1)}{n}F_\alpha(f_1,f_2)},$$

где

$$s^2 = \frac{1}{k\,(n-1)}\sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij}-\bar{x})^2; \quad \bar{x} = \frac{1}{k}\sum_{j=1}^{k}\bar{x}_j, \quad \bar{x}_j = \frac{1}{n}x_{ij},$$

$F_\alpha(f_1,f_2)$ — α-квантиль распределения Фишера с $f_1 = k-1$ и $f_2 = k(n-1)$ степенями свободы (может быть взята из таблиц или аппроксимирована — см. раздел 1.1.10). Критерий Шеффе является грубым критерием и особенно пригоден тогда, когда имеется подозрение о неравенстве дисперсий s_j^2 $(j = 1, \ldots, k)$ между собой.

Задача 196. *В условиях задачи* 190 *проверить нулевую гипотезу равенства средних критерием Шеффе при доверительной вероятности* $\alpha = 0{,}95$.

Имеем

$$n = 8; \quad k = 5; \quad \bar{x}_1 = 26{,}375; \quad \bar{x}_2 = 29{,}25; \quad \bar{x}_3 = 27{,}125; \quad \bar{x}_4 = 24{,}625; \quad \bar{x}_5 = 47{,}75;$$
$$s = 16{,}563; \quad f_1 = k-1 = 4; \quad f_2 = k\cdot(n-1) = 35.$$

Из таблиц F-распределения для $\alpha = 0{,}95$, $f_1 = 4$ и $f_2 = 35$ находим $F_{0,95}(4;35) = 2{,}65$. Следовательно, $D_1 = 16{,}563\cdot\sqrt{\frac{2\cdot 4}{8}\cdot 2{,}65} = 26{,}963$.

Имеем $\bar{x}_5 - \bar{x}_1 = 47{,}75 - 24{,}625 = 23{,}125 < D_1 = 26{,}963$.

Очевидно, что дальнейшая проверка не имеет смысла, и критерий Шеффе не отклоняет нулевую гипотезу.

4.1.1.2.8. Критерий Стьюдента–Ньюмена–Кейлса

Критерий предложен в [402, 403]. Алгоритм его применения аналогичен критерию Шеффе (см. раздел 4.1.1.2.7), отличаясь от него только критическим значением разности средних.

В критерии Стьюдента–Ньюмена–Кейлса критическая разность средних равна

$$D_2 = \frac{s}{\sqrt{n}}q\,(k^*, f, \alpha)\,,$$

где $q(k^*, f, \alpha)$ — множитель, значения которого приведены в табл. 127; k^* — количество средних значений между сравниваемыми, включая сами эти значения (например, для разности $\bar{x}_5 - \bar{x}_4$ имеем $k^* = 2$, а для разности $\bar{x}_5 - \bar{x}_2$ имеем $k^* = 4$); $f = k(n-1)$ — число степеней свободы при оценке s^2; $s^2 = \dfrac{1}{k(n-1)}\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{n}(x_{ij}-\bar{x})^2$;

$$\bar{x} = \frac{1}{kn}\sum_{j=1}^{k}\sum_{i=1}^{n}x_{ij}.$$

Задача 197. *В условиях задачи* 190 *проверить гипотезу равенства средних критерием Стьюдента–Ньюмена–Кейлса при доверительной вероятности* $\alpha = 0{,}95$.

Из табл. 127 для $\bar{x}_5 - \bar{x}_1 (k^* = 5)$, $f = k\cdot(n-1) = 35$ и $\alpha = 0{,}95$ имеем $q(5, 35, 0{,}95) = 4{,}07$. Тогда $D_2 = q\cdot\frac{s}{\sqrt{n}} = \frac{4{,}07\cdot 16{,}563}{\sqrt{8}} = 23{,}833$.

Так как $\bar{x}_5 - \bar{x}_1 = 47{,}75 - 24{,}625 = 23{,}125 \approx D_2 = 23{,}833$, нулевая гипотеза отклоняется.

4.1.1.2.9. Критерий Дункана

Дункан [404] предложил модификацию критерия Стьюдента–Ньюмена–Кейлса, заменив уровень значимости α на $\alpha' = 1 - (1-\alpha)^{k-1}$.

Метод Дункана дает наименьшие допуски для разности средних и наиболее широко применим. Он позволяет „уловить" различие между средними, которые „пропускает" критерий Стьюдента–Ньюмена–Кейлса. Алгоритм применения критерия Дункана D_3 идентичен алгоритму критерия Стьюдента–Ньюмена–Кейлса (см. раздел 4.1.1.2.8), отличаясь от него только использованием вместо множителя $q(k^*, f, \alpha)$ множителя $q(k^*, f, \alpha')$, значения которого для $\alpha = 0{,}95$ приведены в табл. 128.

Задача 198. *В условиях задачи* 190 *проверить гипотезу равенства средних критерием Дункана при доверительной вероятности* $\alpha = 0{,}95$.

Из табл. 128 для $\bar{x}_5 - \bar{x}_1$ $(k^* = 5)$ имеем при $f = k\cdot(n-1) = 35$ $q(5;35;0{,}95) = 3{,}18$. Тогда $D_3 = q\cdot\dfrac{s}{\sqrt{n}} = \dfrac{3{,}18\cdot 16{,}563}{\sqrt{8}} = 18{,}622$.

Так как $\bar{x}_5 - \bar{x}_1 = 47{,}75 - 24{,}625 = 23{,}125 > D_3 = 18{,}622$, нулевая гипотеза отклоняется ($\bar{x}_5$ признается с вероятностью 0,95 бóльшим, чем $\bar{x}_1$).

Для $\bar{x}_5 - \bar{x}_2$ $(k^* = 4)$ имеем $q(4;35;0{,}95) = 3{,}12$ и $D_3 = \dfrac{3{,}12\cdot 16{,}563}{\sqrt{8}} = 18{,}270$.

Так как $\bar{x}_5 - \bar{x}_2 = 47{,}75 - 26{,}375 = 21{,}375$, то и разница между $\bar{x}_5$ и $\bar{x}_2$ признается значимой. Далее по аналогии для

$$\bar{x}_5 - \bar{x}_3 \quad (k^* = 3) \quad \text{имеем} \quad D_3 = \frac{3,02\cdot 16,563}{\sqrt{8}} = 17,684.$$

Имеем $\bar{x}_5 - \bar{x}_3 = 47{,}74 - 27{,}125 = 20{,}625 > D_3$, что также позволяет прийти к заключению о значимости разности между $\bar{x}_5$ и $\bar{x}_3$.

Для $\bar{x}_5 - \bar{x}_3$ имеем $D_3 = 16{,}748$ и $\bar{x}_5 - \bar{x}_4 = 47{,}75 - 29{,}75 = 18{,}0 > D_3 = 16{,}748$.

Таким образом, $\bar{x}_5$ значимо больше, чем $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$ и $\bar{x}_4$. Можно показать, что остальные пары средних критерием Дункана статистически не различимы.

Отметим, что, в отличие от ряда ранее использовавшихся критериев, критерий Дункана, как более чувствительный критерий, отклонил нулевую гипотезу достаточно уверенно.

4.1.1.2.10. Критерий Линка–Уоллеса

Критерий предложен в [405, 406] и достаточно полно изложен в [9]. Его применение предполагает нормальность распределения x_{ij}, равенство дисперсий s_j^2 и равные объемы сравниваемых выборок (n). Обозначим через $R_j = \max\limits_{1\leqslant i\leqslant n} x_{ij} - \min\limits_{1\leqslant i\leqslant n} x_{ij}$ размах j-й выборки, а через $R = \max\limits_{1\leqslant j\leqslant k} \bar{x}_j - \min\limits_{1\leqslant j\leqslant k} \bar{x}_j$ размах средних k выборок.

Нулевая гипотеза равенства средних отклоняется, если

$$K = \frac{nR}{\sum\limits_{j=1}^{k} R_j} > K(n,k,\alpha),$$

где $K(n,k,\alpha)$ — критические значения, приведенные в табл. 129.

Если $\bar{x}_i - \bar{x}_j > K(n,k,\alpha)\dfrac{\sum\limits_{j=1}^{k} R_j}{n}$, то с вероятностью α средние $\bar{x}_i$ и $\bar{x}_j$ признаются статистически различимыми.

Таблица 127

Множители $q(k^*, f, \alpha)$ для критерия Стьюдента–Ньюмена–Кейлса при доверительной вероятности $\alpha = 0{,}95$ [9]

f	k^*													
	2	3	4	5	6	7	8	9	10	12	14	16	18	20
3	4,50	5,91	6,82	7,50	8,04	8,48	8,85	9,18	9,46	9,95	10,35	10,69	10,98	11,24
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83	8,21	8,52	8,79	9,03	9,23
5	3,63	4,60	5,22	5,67	6,03	6,33	6,58	6,80	6,99	7,32	7,60	7,83	8,03	8,21
6	3,46	4,34	4,90	5,30	5,63	5,90	6,12	6,32	6,49	6,79	7,03	7,24	7,43	7,59
7	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,43	6,66	6,85	7,02	7,17
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,18	6,39	6,57	6,73	6,87
9	3,20	3,95	4,41	4,76	5,02	5,24	5,43	5,59	5,74	5,98	6,19	6,36	6,51	6,64
10	3,15	3,88	4,33	4,65	4,91	5,12	5,30	5,46	5,60	5,83	6,03	6,19	6,34	6,47
11	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49	5,71	5,90	6,06	6,20	6,33
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,39	5,61	5,80	5,95	6,09	6,21
13	3,05	3,73	4,15	4,45	4,69	4,88	5,05	5,19	5,32	5,53	5,71	5,86	5,99	6,11
14	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,46	5,64	5,79	5,91	6,03
15	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,20	5,40	5,57	5,72	5,85	5,96
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15	5,35	5,52	5,66	5,79	5,90
17	2,98	3,63	4,02	4,30	4,52	4,70	4,86	4,99	5,11	5,31	5,47	5,61	5,73	5,84
18	2,97	3,61	4,00	4,28	4,49	4,67	4,82	4,96	5,07	5,27	5,43	5,57	5,69	5,79
19	2,96	3,59	3,98	4,25	4,47	4,64	4,79	4,92	5,04	5,23	5,39	5,53	5,65	5,75
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01	5,20	5,36	5,49	5,61	5,71
22	2,93	3,55	3,93	4,20	4,41	4,58	4,72	4,85	4,96	5,15	5,30	5,43	5,56	5,64
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92	5,10	5,25	5,38	5,49	5,59
26	2,91	3,51	3,88	4,14	4,34	4,51	4,65	4,78	4,89	5,06	5,21	5,34	5,46	5,55
28	2,90	3,50	3,86	4,12	4,32	4,48	4,62	4,75	4,86	5,03	5,18	5,30	5,42	5,51
30	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	5,00	5,15	5,27	5,38	5,47
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,73	4,90	5,04	5,16	5,27	5,36
50	2,84	3,41	3,76	4,00	4,19	4,34	4,47	4,58	4,69	4,85	4,99	5,10	5,20	5,29
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,81	4,94	5,06	5,15	5,24
120	2,80	3,36	3,68	3,92	4,10	4,24	4,36	4,47	4,56	4,74	4,84	4,95	5,04	5,13

Таблица 128

Множители $q(k^*, f, \alpha')$ для критерия Дункана при доверительной вероятности $\alpha = 0{,}95$ [119]

f	k^*													
	2	3	4	5	6	7	8	9	10	12	14	16	18	20
3	4,50	4,52	4,52	4,52	4,52	4,52	4,52	4,52	4,52	4,52	4,52	4,52	4,52	4,52
4	3,93	4,01	4,03	4,03	4,03	4,03	4,03	4,03	4,03	4,03	4,03	4,03	4,03	4,03
5	3,64	3,75	3,80	3,81	3,81	3,81	3,81	3,81	3,81	3,81	3,81	3,81	3,81	3,81
6	3,46	3,59	3,65	3,68	3,69	3,70	3,70	3,70	3,70	3,70	3,70	3,70	3,80	3,80
7	3,34	3,48	3,55	3,59	3,61	3,62	3,63	3,63	3,63	3,63	3,63	3,63	3,63	3,63
8	3,26	3,40	3,48	3,52	3,55	3,57	3,58	3,58	3,58	3,58	3,58	3,58	3,58	3,58
9	3,20	3,34	3,42	3,47	3,50	3,52	3,54	3,54	3,55	3,55	3,55	3,55	3,55	3,55
10	3,15	3,29	3,38	3,43	3,46	3,49	3,50	3,52	3,52	3,53	3,53	3,53	3,53	3,53
11	3,11	3,26	3,34	3,40	3,44	3,46	3,48	3,49	3,50	3,51	3,51	3,51	3,51	3,51
12	3,08	3,22	3,31	3,37	3,41	3,44	3,46	3,47	3,48	3,50	3,50	3,50	3,50	3,50
13	3,06	3,20	3,29	3,35	3,39	3,42	3,44	3,46	3,47	3,48	3,49	3,49	3,49	3,49
14	3,03	3,18	3,27	3,33	3,37	3,40	3,43	3,44	3,46	3,47	3,48	3,48	3,48	3,48
15	3,01	3,16	3,25	3,31	3,36	3,39	3,41	3,43	3,45	3,46	3,48	3,48	3,48	3,48
16	3,00	3,14	3,24	3,30	3,34	3,38	3,40	3,42	3,44	3,46	3,47	3,48	3,48	3,48
17	2,98	3,13	3,22	3,28	3,33	3,37	3,39	3,41	3,43	3,45	3,46	3,47	3,48	3,48
18	2,97	3,12	3,21	3,27	3,32	3,36	3,38	3,40	3,42	3,44	3,46	3,47	3,47	3,47
19	2,96	3,11	3,20	3,26	3,31	3,35	3,38	3,40	3,42	3,44	3,46	3,47	3,47	3,47
20	2,95	3,10	3,19	3,24	3,30	3,34	3,37	3,39	3,41	3,44	3,45	3,46	3,47	3,47
24	2,92	3,07	3,16	3,23	3,28	3,32	3,34	3,37	3,39	3,42	3,44	3,46	3,46	3,47
30	2,89	3,04	3,13	3,20	3,25	3,29	3,32	3,35	3,37	3,40	3,43	3,45	3,46	3,47
40	2,86	3,01	3,10	3,17	3,22	3,27	3,30	3,33	3,35	3,39	3,42	3,44	3,46	3,47
60	2,83	2,98	3,07	3,14	3,20	3,24	3,28	3,31	3,33	3,37	3,41	3,43	3,45	3,47
120	2,80	2,95	3,04	3,12	3,17	3,22	3,25	3,29	3,31	3,36	3,39	3,42	3,45	3,47

Таблица 129

Критические значения $K(n, k, \alpha)$ для критерия Линка–Уоллеса при доверительной вероятности $\alpha = 0{,}95$

n	k													
	2	3	4	5	6	7	8	9	10	12	14	16	18	20
2	1,90	1,44	1,14	0,94	0,80	0,70	0,62	0,56	0,51	0,43	0,38	0,33	0,30	0,27
4	1,62	1,25	1,01	0,84	0,72	0,63	0,57	0,51	0,47	0,40	0,35	0,31	0,28	0,25
5	1,53	1,19	0,96	0,82	0,70	0,61	0,55	0,50	0,45	0,39	0,34	0,30	0,27	0,25
6	1,50	1,17	0,95	0,80	0,69	0,61	0,55	0,40	0,45	0,39	0,34	0,30	0,27	0,26
7	1,49	1,17	0,95	0,80	0,69	0,61	0,55	0,50	0,45	0,39	0,34	0,30	0,27	0,25
8	1,49	1,18	0,96	0,81	0,70	0,62	0,55	0,50	0,46	0,39	0,35	0,31	0,28	0,25
9	1,50	1,19	0,97	0,82	0,71	0,62	0,56	0,51	0,47	0,40	0,35	0,31	0,28	0,25
10	1,52	1,20	0,98	0,83	0,72	0,63	0,57	0,52	0,47	0,41	0,36	0,32	0,29	0,26
11	1,54	1,22	0,99	0,84	0,73	0,64	0,58	0,52	0,48	0,41	0,36	0,32	0,29	0,27
12	1,56	1,23	1,01	0,85	0,74	0,65	0,58	0,53	0,49	0,42	0,37	0,33	0,30	0,27
13	1,58	1,25	1,02	0,86	0,75	0,66	0,59	0,54	0,49	0,42	0,37	0,33	0,30	0,27
14	1,60	1,26	1,03	0,87	0,76	0,67	0,60	0,55	0,50	0,43	0,38	0,34	0,30	0,28
15	1,62	1,28	1,05	0,89	0,77	0,68	0,61	0,55	0,51	0,44	0,38	0,34	0,31	0,28
16	1,64	1,30	1,06	0,90	0,78	0,69	0,62	0,56	0,52	0,44	0,39	0,35	0,31	0,29
17	1,66	1,31	1,08	0,91	0,79	0,70	0,63	0,57	0,52	0,45	0,39	0,35	0,32	0,29
18	1,68	1,33	1,09	0,92	0,80	0,71	0,64	0,58	0,53	0,46	0,40	0,36	0,32	0,29
19	1,70	1,35	1,10	0,93	0,81	0,72	0,64	0,59	0,54	0,46	0,41	0,36	0,33	0,30
20	1,72	1,36	1,12	0,95	0,82	0,73	0,65	0,59	0,54	0,47	0,41	0,37	0,33	0,30
30	1,92	1,52	1,24	1,05	0,01	0,81	0,73	0,66	0,60	0,52	0,46	0,41	0,37	0,34
40	2,08	1,66	1,35	1,14	0,99	0,88	0,79	0,72	0,66	0,57	0,50	0,44	0,40	0,37
50	2,23	1,77	1,45	1,22	1,06	0,94	0,85	0,77	0,71	0,61	0,53	0,48	0,43	0,39
100	2,81	2,23	1,83	1,55	1,34	1,19	1,07	0,97	0,89	0,77	0,67	0,60	0,55	0,50

Задача 199. *В условиях задачи 190 проверить гипотезу равенства средних критерием Линка–Уоллеса при доверительной вероятности $\alpha = 0{,}95$.*

Имеем

$$R_1 = 45 - 11 = 34; \quad R_2 = 52 - 15 = 37; \quad R_3 = 58 - 8 = 50; \quad R_4 = 61 - 4 = 57;$$

$$R_5 = 70 - 24 = 46; \quad \sum_{j=1}^{5} R_j = 224; \quad R = \max \bar{x}_j - \min \bar{x}_j = 47{,}75 - 24{,}625 = 23{,}125.$$

Вычисляем

$$K = \frac{8 \cdot 23{,}125}{224} = 0{,}826.$$

Из табл. 129 для $\alpha = 0{,}95$, $k = 5$ и $n = 8$ имеем $K(8, 5, 0{,}95) = 0{,}81$.

Так как $K = 0{,}826 > K(8, 5, 0{,}95) = 0{,}81$, нулевая гипотеза отклоняется.

Находим критическую разницу средних

$$\frac{K(n, k, \alpha)}{n} \cdot \sum_{j=1}^{k} R_j = \frac{0{,}81 \cdot 224}{8} = 22{,}68.$$

Видим, что $\bar{x}_5 - \bar{x}_1 = 47{,}75 - 24{,}625 = 23{,}125$ превосходит эту величину.

Следовательно, разность между $\bar{x}_5$ и $\bar{x}_1$ признается значимой.

4.1.1.3. Сравнение двух дисперсий

Для двух выборок нормально распределенных случайных величин $x_1, \ldots, x_n$ и $y_1, \ldots, y_m$ необходимо проверить гипотезу равенства дисперсий σ_1^2 и σ_2^2, опираясь на их выборочные оценки s_1^2 и s_2^2.

4.1.1.3.1. Критерий Фишера

Если выборочными оценками максимального правдоподобия дисперсий являются

$$s_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{и} \quad s_2^2 = \frac{1}{m-1} \sum_{i=1}^{m} (y_i - \bar{y})^2,$$

то статистика критерия Фишера записывается как $F = \dfrac{s_1^2}{s_2^2}$.

При справедливости нулевой гипотезы $H_0\colon \sigma_1^2 = \sigma_2^2$ статистика критерия имеет распределение Фишера (см. раздел 1.1.10) с $f_1 = n - 1$ и $f_2 = m - 1$ степенями свободы, где n и m — объемы сравниваемых выборок.

Если $F > F_{\frac{1+\alpha}{2}}(n-1; m-1)$ и $F < F_{\frac{1-\alpha}{2}}(n-1; m-1)$, то нулевая гипотеза отклоняется в пользу альтернативы $H_1\colon \sigma_1^2 \neq \sigma_2^2$.

Если $F > F_\alpha(n-1; m-1)$, то нулевая гипотеза отклоняется в пользу альтернативы $H_1'\colon \sigma_1^2 > \sigma_2^2$ (α — доверительная вероятность). В числителе всегда должна стоять бóльшая по величине из двух сравниваемых дисперсий.

Критерий Фишера очень чувствителен к отклонениям от нормальности [407–409] распределения x_i, y_i. Его устойчивость к отклонениям от нормальности может быть повышена соответствующей корректировкой степеней свободы [87]. Вместо f_1 и f_2 в этом случае используются степени свободы

$$f_1' = df \quad \text{и} \quad f_2' = bf, \quad \text{где}$$

$$d = \left\{1 + \frac{1}{2}\left(\frac{n+m-4}{n+m-(b_2-3)}\right)(b_2 - 3)\right\}^{-1} b = (m+n)\,\frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^4 + \sum\limits_{i=1}^{m} (y_i - \bar{y})^4}{\left\{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2 + \sum\limits_{i=1}^{m} (y_i - \bar{y})^2\right\}^2}.$$

В дальнейшем процедура проверки нулевой гипотезы не отличается от обычного F-критерия. Критические значения F-статистики приведены в таблицах (ссылки на таблицы см. в разделе 1.1.10).

Задача 200. *Имеются две выборки нормально распределенных случайных величин* ($n = m = 10$):

x_i:	2,1;	3,1;	4,8;	6,1;	7,4;	8,5;	10,1;	12,1;	14,0;	15,6;
y_i:	4,6;	6,1;	8,2;	9,8;	9,9;	10,4;	13,1;	14,5;	16,1;	19,1.

Необходимо проверить гипотезу равенства дисперсий $H_0: \sigma_1^2 = \sigma_2^2$ *против альтернативы* $H_1: \sigma_1^2 \neq \sigma_2^2$ *при доверительной вероятности* $\alpha = 0{,}95$.

Имеем

$$\bar{x} = 8{,}38; \quad \bar{y} = 11{,}18;$$

$$s_1^2 = \frac{1}{10-1} \cdot \sum_{i=1}^{10} (x_i - \bar{x})^2 = 20{,}757; \quad s_2^2 = \frac{1}{10-1} \cdot \sum_{i=1}^{10} (y_i - \bar{y})^2 = 20{,}419.$$

Далее $F = \dfrac{s_1^2}{s_2^2} = \dfrac{20{,}757}{20{,}419} = 1{,}016$. Из таблиц находим

$$F_{\frac{1+\alpha}{2}}(n-1; m-1) = F_{\frac{1+0{,}95}{2}}(10-1; 10-1) = F_{0{,}975}(9;9) = 4{,}03.$$

Так как $F = 1{,}016 < F_{0{,}975}(9{,}9) = 4{,}03$, нулевая гипотеза не отклоняется.

Рассмотрим теперь критерий со скорректированными степенями свободы. Имеем

$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 186{,}813; \quad \sum_{i=1}^{10} (y_i - \bar{y})^2 = 183{,}771; \quad \sum_{i=1}^{10} (x_i - x)^4 = 6439{,}996;$$

$$\sum_{i=1}^{10} (y_i - \bar{y})^4 = 7281{,}725; \quad b_2 = (10+10) \cdot \frac{6439{,}996 + 7281{,}725}{(186{,}813 + 183{,}771)^2} = 1{,}99832;$$

$$d = \left\{1 + 0{,}5 \cdot \frac{10+10-4}{10+10-3-(1{,}99832-3)} \cdot (1{,}99832-3)\right\}^{-1} = 1{,}802.$$

Окончательно имеем $f_1' = d \cdot f_1 = 1{,}802 \cdot 9 = 16{,}22$ и $f_2' = d \cdot f_2 = 16{,}22$.

Из таблиц [87] для дробных степеней свободы и $\dfrac{1+\alpha}{2} = 0{,}975$ имеем $F_{0{,}975}(16{,}22; 16{,}22) \approx 2{,}65$. Так как $F = 1{,}016 < F_{0{,}975} = 2{,}65$, нулевая гипотеза не отклоняется и в этом случае.

4.1.1.3.2. Критерий Романовского [207]

Статистика критерия:

$$R = \frac{|Q-1|}{\sigma_Q},$$

где $Q = \dfrac{m-3}{m-1}\dfrac{s_1^2}{s_2^2}$; $\sigma_Q = \sqrt{\dfrac{2(n+m-4)}{(n-1)(m-5)}}$.

Если $R \geqslant 3$, то нулевая гипотеза равенства дисперсий отклоняется с достоверностью не менее 0,89.

Задача 201. *Для данных задачи* 200 *проверить гипотезу равенства дисперсий критерием Романовского.*

Имеем

$$Q = \frac{10-3}{10-1} \cdot \frac{20{,}757}{20{,}419} = 0{,}7906; \quad \sigma_Q = \sqrt{\frac{2 \cdot (10+10-4)}{(10-1) \cdot (10-5)}} = 0{,}84327; \quad R = \frac{|0{,}7906 - 1|}{0{,}84327} = 0{,}248.$$

Так как $R < 3$, нулевая гипотеза равенства дисперсий не отклоняется.

4.1.1.3.3. Критерий отношения размахов

Статистика критерия имеет вид

$$F^* = \frac{\omega_n}{\omega_m},$$

где $\omega_n = x_{\max} - x_{\min}$; $\omega_m = y_{\max} - y_{\min}$ — размахи сравниваемых выборок.

Если $F^* > F^*_\alpha$, где F^*_α — критическое значение статистики, то нулевая гипотеза отклоняется с вероятностью α. Критические значения $F^*_\alpha(n, m)$ приведены в табл. 130, заимствованной из [25]. При $n \leqslant 15$ мощность критерия отношения размахов практически не отличается от мощности критерия Фишера (см. раздел 4.1.1.3.1) (в числителе всегда должно быть наибольшее из двух значение размаха).

Таблица 130

Критические значения $F^*(n, m)$ отношения размахов
(α — доверительная вероятность) [25]

n	m	α		n	m	α		n	m	α	
		0,95	0,99			0,95	0,99			0,95	0,99
3	2	3,90	7,37	8	6	1,85	2,57	11	10	1,77	2,30
	3	4,37	9,99		7	1,94	2,67		11	1,81	2,34
	4	5,14	11,71		8	2,01	2,76		12	1,84	2,39
	5	5,71	12,98		9	2,08	2,84		13	1,88	2,42
4	3	2,03	3,72		10	2,13	2,91		14	1,91	2,46
	4	2,66	4,79		11	2,18	2,98		15	1,93	2,49
	5	3,07	5,49		12	2,23	2,56	12	10	1,69	2,18
	6	3,38	6,01	9	5	1,63	2,23		11	1,73	2,22
5	3	1,60	2,66		6	1,72	2,34		12	1,77	2,26
	4	2,06	3,32		7	1,80	2,43		13	1,80	2,30
	5	2,35	3,76		8	1,87	2,14		14	1,82	2,33
	6	2,57	4,08		9	1,93	2,20		15	1,85	2,36
	7	2,74	4,36		10	1,98	2,26	13	10	1,63	2,08
	8	3,00	4,73		11	2,02	2,31		11	1,67	2,12
6	4	1,99	2,98		12	2,07	2,75		12	1,70	2,16
	5	2,16	3,22		13	2,10	2,80		13	1,73	2,19
	6	2,30	3,41		14	2,14	2,84		14	1,76	2,22
	7	2,41	3,56		15	2,17	2,88		15	1,78	2,25
	8	2,51	3,70	10	6	1,63	2,17	14	10	1,58	2,00
	9	2,59	3,81		7	1,70	2,26		11	1,61	2,04
	10	2,67	3,92		8	1,76	2,33		12	1,64	2,07
7	5	1,92	2,74		9	1,81	2,39		13	1,67	2,10
	6	2,03	2,89		10	1,86	2,45		14	1,70	2,13
	7	2,13	3,02		11	1,90	2,50		15	1,72	2,16
	8	2,21	3,13		12	1,94	2,54	15	10	1,53	1,93
	9	2,28	3,22		13	1,98	2,59		11	1,57	1,97
	10	2,35	3,30		14	2,01	2,62		12	1,60	2,00
	11	2,40	3,38		15	2,04	2,66		13	1,62	2,03
	12	2,46	3,44	11	8	1,67	2,19		14	1,65	2,06
8	5	1,75	2,44		9	1,72	2,25		15	1,67	2,08

Задача 202. *Для данных задачи 200 проверить гипотезу равенства дисперсий критерием отношения размахов при доверительной вероятности* $\alpha = 0{,}95$.

Имеем $\omega_n = 15{,}6 - 2{,}1 = 13{,}5$ и $\omega_m = 19{,}1 - 4{,}6 = 14{,}5$.

Далее $F^* = \dfrac{14{,}5}{13{,}5} = 1{,}074$.

Из табл. 130 для $n = m = 10$ и $\alpha = 0{,}95$ имеем $F^*_{0,95}(10{,}10) = 1{,}86$.

Так как $F^* = 1{,}074 < F^*_{0,95}(10{,}10) = 1{,}86$, нулевая гипотеза равенства дисперсий не отклоняется.

4.1.1.3.4. Критерий „стьюдентизированного" размаха

Статистика критерия:

$$q = \frac{\omega_n}{s_m}, \quad \text{где} \quad \omega_n = x_{\max} - x_{\min}; \quad s_m = \left\{ \frac{1}{m-1} \sum_{i=1}^{m} (y_i - \bar{y}) \right\}^{\frac{1}{2}}.$$

Нулевая гипотеза равенства дисперсий отклоняется, если $q > q_\alpha$, где q_α — критическое значение статистики, приведенное в табл. 123 (следует помнить, что здесь $f = m - 1$).

Задача 203. *Для данных задачи 200 проверить гипотезу равенства дисперсий критерием „стьюдентизированного" размаха при $\alpha = 0{,}95$.*

Имеем

$$\omega_n = 15{,}6 - 2{,}1 = 13{,}5; \quad \bar{y} = 11{,}18; \quad s_m = \left\{ \frac{1}{10-1} \cdot \sum_{i=1}^{10} (y_i - \bar{y})^2 \right\}^{\frac{1}{2}} = 4{,}519.$$

Тогда $q = \dfrac{13{,}5}{4{,}519} = 2{,}99$.

Из табл. 123 для $n = 10$, $f = 10 - 1 = 9$ и $\alpha = 0{,}95$ имеем $q_{0,95} \approx 5{,}5$.

Так как $q = 2{,}99 < q_{0,95} = 5{,}5$, нулевая гипотеза равенства дисперсий не отклоняется.

4.1.1.3.5. Критерий Аризоно–Охты

Предложен в [410] и основан на энтропийном критерии нормальности Васичека (см. раздел 3.2.2.2).

Если обе выборки упорядочены:

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n \quad \text{и} \quad y_1 \leqslant y_2 \leqslant \ldots \leqslant y_m,$$

то статистика записывается в форме

$$K = \frac{c}{d} \cdot \frac{n}{m} \cdot \frac{\left\{ \prod\limits_{i=1}^{n} (x_{i+d} - x_{i-d}) \right\}^{\frac{1}{n}}}{\left\{ \prod\limits_{j=1}^{m} (y_{j+c} - y_{j-c}) \right\}^{\frac{1}{m}}},$$

где x_i, y_j — порядковые статистики выборок (т. е. упорядоченные по величине элементы выборок).

При $i, j \leqslant 1$ $x_i\ (y_j) = x_1\ (y_1)$ и при $i\ (j) > n\ (m)$ $x_i\ (y_j) = x_n\ (y_m)$.

Критерий практически не уступает по мощности критерию Фишера (см. раздел 4.1.1.3.1) при $n\ (m) \leqslant 20$.

При $K_1(\alpha) \leqslant K \leqslant K_2(\alpha)$ нулевая гипотеза равенства дисперсий не отклоняется. Здесь $K_1(\alpha)$ и $K_2(\alpha)$ — критические значения статистики, приведенные в табл. 131. При $K < K_1(\alpha)$ или $K > K_2(\alpha)$ нулевая гипотеза отклоняется (здесь α — доверительная вероятность).

Таблица 131

Критические значения $K_1(\alpha)$ (верхняя строка) и $K_2(\alpha)$ (нижняя строка) статистики Аризоно–Охты для доверительной вероятности $\alpha = 0{,}95$ [410]

n	c	$m = 3$	5	8	10	15	20
		$d = 1$	2	3	3	3	3
3	1	0,139					
		7,425					
5	2	0,105	0,316				
		2,613	3,087				
8	3	0,087	0,264	0,435			
		1,892	2,045	2,316			
10	3	0,082	0,244	0,398	0,468		
		1,619	1,739	1,945	2,101		
15	3	0,072	0,218	0,360	0,425	0,559	
		1,270	1,350	1,488	1,617	1,804	
20	3	0,069	0,199	0,335	0,408	0,530	0,602
		1,167	1,205	1,346	1,427	1,576	1,692

В табл. 131 указаны соотношения $n(m)$ и $c(d)$, обеспечивающие максимальную мощность критерия.

Задача 204. *В условиях задачи 200 проверить нулевую гипотезу равенства дисперсий критерием Аризоно–Охты.*

Имеем $n = m = 10$. Из табл. 131 выбираем $c = d = 3$.

Вычисляем (имея в виду, что $x_i = x_1$ при $i \leqslant 1$ и $x_i = x_n$ при $i \geqslant n$):

$$\prod_{i=1}^{10}(x_{i+3} - x_{i-3}) = (x_4 - x_1)\cdot(x_5 - x_1)\cdots(x_{10} - x_6)\cdot(x_{10} - x_7) =$$

$$= (6{,}2 - 2{,}1)\cdot(7{,}4 - 2{,}1)\cdots(15{,}6 - 8{,}5)\cdot(15{,}6 - 10{,}1) = 2{,}80223251\cdot 10^8;$$

$$\left(2{,}8023251\cdot 10^8\right)^{\frac{1}{10}} = 6{,}994223;$$

$$\prod_{j=1}^{10}(y_{j+3} - y_{j-3}) = (9{,}8 - 4{,}6)\cdot(9{,}9 - 4{,}6)\cdots(19{,}1 - 10{,}4)\cdot(19{,}1 - 13{,}1) = 43770949;$$

$$(43770949)^{\frac{1}{10}} = 5{,}8092303.$$

Окончательно получаем $K = \dfrac{3}{3}\cdot\dfrac{10}{10}\cdot\dfrac{6{,}9944223}{5{,}8092303} = 1{,}204$.

Для $n = m = 10$, $c = d = 3$ и $\alpha = 0{,}95$ из табл. 131 имеем

$$K_1(0{,}95) = 0{,}468 \quad \text{и} \quad K_2(0{,}95) = 2{,}101.$$

Так как $K_1(0{,}95) = 0{,}468 < K = 1{,}204 < K_2(0{,}95) = 2{,}101$, нулевая гипотеза равенства дисперсий не отклоняется.

4.1.1.4. Сравнение нескольких ($k > 2$) дисперсий

Пусть $s_1^2, s_2^2, \ldots, s_k^2$ — взаимно независимые выборочные оценки дисперсий $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ по выборкам объема $n_1, n_2, \ldots, n_k$. Проверяется нулевая гипотеза $H_0\colon \sigma_1^2 = \sigma_{i+1}^2$ против альтернативы $H_0\colon \sigma_1^2 \neq \sigma_{i+1}^2$ (для $i = 1, 2, \ldots, k-1$).

4.1.1.4.1. Критерий Бартлетта

Предложен в [411], основан на статистике

$$M = n\ln\left\{\frac{1}{n}\sum_{i=1}^{k}(i-1)\,s_i^2\right\} - \sum_{i=1}^{k}(n_i-1)\ln s_i^2, \quad \text{где} \quad n = \sum_{i=1}^{k}(n_i-1).$$

При $n_i > 3$ $(i = 1, \ldots, k)$ и справедливости нулевой гипотезы величина $B = M\left\{1 + \frac{1}{3\,(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{n}\right)\right\}^{-1}$ имеет распределение χ^2 с $f = k-1$ степенями свободы (о распределении χ^2 см. раздел 1.1.8).

Если $B > \chi^2_\alpha(f)$, то с достоверностью α нулевая гипотеза отклоняется. Критические значения $\chi^2_\alpha(f)$ содержатся в табл. 55 или могут быть вычислены с помощью аппроксимаций.

При $f > 30$ можно использовать аппроксимацию Вилсона–Хилферти (см. аппроксимацию 3 в разделе 1.1.8) $\chi^2_\alpha(f) = f\left(1 + \frac{2}{9f} + u_\alpha\sqrt{\frac{2}{9f}}\right)^3$, где u_α — квантиль стандартного нормального распределения. Критерий очень чувствителен к отклонениям от нормальности распределения исследуемых случайных величин. Если нет уверенности в нормальности распределения (критерии проверки нормальности распределения рассмотрены в разделе 3.2), им не рекомендуется пользоваться. При отклонении от нормальности рекомендуется вместо статистики M пользоваться ее модификацией [22, 87, 412]

$$M^* = \frac{f_2 M}{f_1(b-M)},$$

где $f_1 = k-1$; $f_2 = \frac{k+1}{(c-1)^2}$; $b = \frac{f_2^2}{f_2(2-c)+c}$; $c = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{m}\frac{1}{n_i-1} - \frac{1}{n}\right)$.

Статистика M^* имеет F-распределение с f_1 и f_2 степенями свободы. Поэтому нулевую гипотезу следует отклонить, если $M^* > F_\alpha\,(f_1, f_2)$.

Задача 205. *Имеются четыре выборки* ($k = 4$) *объема* $n = 5$ *каждая*

$$x_{i1}\colon\ 3,\quad 4,\quad 5,\quad 6,\quad 7; \qquad x_{i2}\colon\ 2,\quad 8,\quad 9,\quad 11,\quad 15;$$
$$x_{i3}\colon\ 9,\quad 11,\quad 15,\quad 20,\quad 28; \qquad x_{i4}\colon\ 4,\quad 6,\quad 8,\quad 10,\quad 16.$$

Необходимо проверить нулевую гипотезу равенства дисперсий в выборках $H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$ *критерием Барлетта при доверительной вероятности* $\alpha = 0{,}95$.

Имеем

$$\bar{x}_1 = 5; \quad \bar{x}_2 = 9; \quad \bar{x}_3 = 16{,}6; \quad \bar{x}_4 = 8{,}8;$$

$$s_1^2 = \frac{1}{5-1}\cdot\sum_{i=1}^{5}(x_{i1}-\bar{x}_1)^2 = 2{,}5 \qquad (s_1 = 1{,}581);$$

$$s_2^2 = \frac{1}{5-1}\cdot\sum_{i=1}^{5}(x_{i2}-\bar{x}_2)^2 = 22{,}5 \qquad (s_2 = 4{,}473);$$

$$s_3^2 = \frac{1}{5-1}\cdot\sum_{i=1}^{5}(x_{i3}-\bar{x}_3)^2 = 58{,}3 \qquad (s_3 = 7{,}635);$$

$$s_4^2 = \frac{1}{5-1}\cdot\sum_{i=1}^{5}(x_{i4}-\bar{x}_4)^2 = 21{,}2 \qquad (s_4 = 4{,}604).$$

Вычисляем

$$n = \sum_{i=1}^{5}(n_i - 1) = 4 \cdot 4 = 16; \quad M = 16 \cdot \ln \frac{1}{16} \cdot \sum_{i=1}^{4} 4 \cdot s_i^2 - \sum_{i=1}^{4} 4 \cdot \ln s_i^2 = 4{,}6088;$$

$$c = 1 + \frac{1}{3 \cdot (4-1)} \cdot \left(\sum_{i=1}^{4} \frac{1}{5-1} - \frac{1}{16}\right) = 1{,}04; \quad B = \frac{M}{c} = \frac{4{,}6088}{1{,}104} = 4{,}474.$$

Из табл. 55 для $f = k - 1 = 4 - 1 = 3$ находим $\chi^2_{0,95}(3) = 7{,}81$.

Так как $B = 4{,}174 < \chi^2_{0,95}(3) = 7{,}81$, нулевая гипотеза не отклоняется. Применим теперь уточненный критерий. Находим

$$f_1 = k - 1 = 3; \quad f_2 = \frac{k+1}{(c-1)^2} = \frac{4+1}{(1{,}104-1)^2} = 462{,}28;$$

$$b = \frac{462{,}28^2}{462{,}28 \cdot (2{,}1{,}104) + 1{,}104} = 514{,}566; \quad M^* = \frac{462{,}28 \cdot 4{,}6088}{3 \cdot (514{,}566 - 4{,}6088)} = 1{,}393.$$

Из таблиц находим $F_{0,95}(3; 462{,}3) = 2{,}60$. Так как $M^* = 1{,}393 < F_{0,95} = 2{,}60$, нулевая гипотеза не отклоняется.

4.1.1.4.2. Критерий Кохрана

Для случая выборок равных объемов ($n_i = n$ при $i = 1, 2, \ldots, k$) Кохран [414] предложил критерий, основанный на статистике

$$g = \frac{\max\limits_{1 \leqslant i \leqslant k} s_i^2}{\sum\limits_{i=1}^{k} s_i^2}.$$

Если $g > g_\alpha(k, n)$, то нулевая гипотеза отклоняется. Значения приведены в табл. 132. Критические значения можно найти также, пользуясь таблицами F-распределения, с помощью соотношения

$$g_\alpha(k, n) = \frac{F_{\frac{k+1-\alpha}{k}}[n-1; (n-1)(k-1)]}{k - 1 + F_{\frac{k-1+\alpha}{k}}[n-1; (n-1)(k-1)]},$$

где $F_\gamma(f_1, f_2)$ — γ-квантиль F-распределения с f_1 и f_2 степенями свободы.

Задача 206. *Проверить нулевую гипотезу равенства дисперсий в условиях задачи* 205 *критерием Кохрана при доверительной вероятности* $\alpha = 0{,}95$.

Имеем $\sum s_i^2 = 2{,}5 + 22{,}5 + 58{,}3 + 21{,}2 = 104{,}5$; $\max s_i^2 = 58{,}3$.

Тогда $g = \dfrac{58{,}3}{104{,}5} = 0{,}558$.

Из табл. 132 для $n = 5$, $k = 4$ и $\alpha = 0{,}95$ имеем $g_{0,95}(4; 5) = 0{,}721$.

Так как $g = 0{,}558 < g_{0,95}(4; 5) = 0{,}721$, нулевая гипотеза не отклоняется. Если воспользоваться таблицами F-распределения, то мы бы получили

$$F_{\frac{4-1+0{,}95}{4}}[(5-1); (5-1) \cdot (4-1)] = F_{0{,}987}(4; 12) \approx 5{,}4; \quad g_{0,95}(4; 5) = \frac{5{,}4}{4 - 1 + 5{,}4} = 0{,}643.$$

Естественно, что приближение, дающее удовлетворительные результаты при $n \geqslant 20$, не обеспечило в нашем случае приемлемой точности.

Таблица 132

Критические значения $g_\alpha(k, n)$ статистики Кохрана для доверительной вероятности $\alpha = 0{,}95$ [25]

k	n					
	2	3	4	5	6	7
2	0,999	0,995	0,979	0,959	0,937	0,917
3	0,993	0,942	0,883	0,833	0,793	0,761
4	0,967	0,864	0,781	0,721	0,676	0,641
5	0,928	0,788	0,696	0,633	0,587	0,553
6	0,883	0,722	0,626	0,563	0,519	0,487
7	0,838	0,664	0,568	0,508	0,466	0,435
8	0,794	0,615	0,521	0,463	0,423	0,393
9	0,754	0,573	0,481	0,425	0,387	0,359
10	0,707	0,536	0,447	0,393	0,357	0,331
12	0,653	0,475	0,392	0,343	0,310	0,286
15	0,548	0,407	0,332	0,288	0,259	0,239
20	0,480	0,330	0,265	0,229	0,205	0,188
30	0,363	0,241	0,191	0,163	0,145	0,133
k	n					
	8	9	10	11	17	37
2	0,899	0,882	0,867	0,854	0,795	0,707
3	0,733	0,711	0,691	0,673	0,606	0,515
4	0,613	0,590	0,570	0,554	0,488	0,406
5	0,526	0,504	0,485	0,470	0,409	0,335
6	0,461	0,440	0,423	0,408	0,353	0,286
7	0,410	0,391	0,375	0,362	0,310	0,249
8	0,370	0,352	0,337	0,325	0,278	0,221
9	0,338	0,321	0,307	0,295	0,251	0,199
10	0,311	0,294	0,281	0,270	0,230	0,181
12	0,268	0,253	0,242	0,232	0,196	0,153
15	0,223	0,210	0,200	0,192	0,161	0,125
20	0,175	0,165	0,157	0,150	0,125	0,096
30	0,123	0,116	0,110	0,105	0,087	0,066

4.1.1.4.3. Критерий Неймана–Пирсона (критерий отношения правдоподобия)

Статистика критерия определяется отношением арифметического среднего всех s_i^2 к их геометрическому среднему (предполагается, что $n_1 = n_2 = \ldots = n_k = n$)

$$H = \frac{\frac{1}{k}\sum_{i=1}^{k} s_i^2}{\left(\prod_{i=1}^{k} s_i^2\right)^{\frac{1}{k}}}.$$

При $H > H_\alpha$ нулевая гипотеза отклоняется. Критические значения критерия приведены в табл. 133. Статистика H применима и для случая, когда средние значения сравниваемых выборок значимо отличаются друг от друга (методы проверки равенства средних см. в разделе 4.1.1).

При равенстве средних $\bar{x}_1 = \ldots = \bar{x}_k$ применима статистика

$$H^* = \frac{1}{nk-1} \sum_{i=1}^{k} x_{ij}.$$

Критические значения H^*_α приведены в табл. 134.

Таблица 133

Критические значения H_α критерия Неймана–Пирсона для доверительной вероятности $\alpha = 0{,}95$

k	n										
	3	4	5	6	7	8	9	10	12	15	20
2	3,21	2,09	1,71	1,52	1,41	1,34	1,29	1,25	1,20	1,15	1,11
3	3,18	2,13	1,74	1,54	1,43	1,35	1,30	1,26	1,21	1,16	1,11
4	3,17	2,08	1,71	1,52	1,41	1,34	1,29	1,25	1,20	1,15	1,11
5	3,05	2,03	1,68	1,50	1,40	1,33	1,28	1,25	1,20	1,15	1,11
10	2,68	1,87	1,58	1,44	1,35	1,29	1,25	1,22	1,17	1,13	1,09
20	2,39	1,75	1,50	1,38	1,30	1,25	1,22	1,19	1,15	1,12	1,08
25	2,31	1,71	1,48	1,36	1,28	1,24	1,20	1,18	1,14	1,11	1,08
50	2,12	1,59	1,40	1,31	1,25	1,21	1,17	1,15	1,12	1,09	1,07

Таблица 134

Критические значения H^*_α критерия Неймана–Пирсона для доверительной вероятности $\alpha = 0{,}95$

k	n										
	3	4	5	6	7	8	9	10	12	15	20
2	5,82	3,05	2,26	1,99	1,78	1,62	1,51	1,41	1,33	1,25	1,17
3	5,73	3,12	2,33	2,06	1,84	1,68	1,54	1,43	1,35	1,26	1,18
4	5,42	3,00	2,26	2,01	1,81	1,66	1,54	1,43	1,35	1,26	1,18
5	5,14	2,92	2,22	1,97	1,78	1,62	1,50	1,42	1,34	1,25	1,18
10	4,36	2,64	2,06	1,84	1,66	1,54	1,44	1,38	1,31	1,23	1,16
20	3,79	2,41	2,93	1,74	1,58	1,44	1,38	1,32	1,26	1,20	1,14
25	3,64	2,38	1,87	1,68	1,54	1,42	1,36	1,31	1,25	1,19	1,13
50	3,29	2,15	1,74	1,58	1,48	1,38	1,32	1,27	1,21	1,17	1,12

Задача 207. *Проверить нулевую гипотезу равенства дисперсий критерием Неймана–Пирсона для данных задачи* 205.

Имеем

$$\bar{x}_1 = 5; \quad \bar{x}_2 = 9; \quad \bar{x}_3 = 16{,}6; \quad \bar{x}_4 = 8{,}8; \quad s_1^2 = 2{,}5; \quad s_2^2 = 22{,}5; \quad s_3^2 = 58{,}3; \quad s_4^2 = 21{,}2.$$

Находим

$$H = \frac{\frac{1}{4} \cdot \sum\limits_{i=1}^{4} s_i^2}{\left(\prod\limits_{i=1}^{4} s_i^2\right)^{\frac{1}{4}}} = \frac{\frac{1}{4} \cdot (2{,}5 + 22{,}5 + 58{,}3 + 21{,}2)}{(2{,}5 \cdot 22{,}5 \cdot 58{,}3 \cdot 21{,}2)^{\frac{1}{4}}} = 1{,}609.$$

Из табл. 133 для $n = 5$, $k = 4$ находим $H_{0,95} = 1{,}71$.

Так как $H = 1{,}609 < H_{0,95} = 1{,}71$, нулевая гипотеза не отклоняется.

4.1.1.4.4. Критерий Блисса–Кохрана–Тьюки

Критерий предложен в [415] как аналог критерия Кохрана (см. раздел 4.1.1.4.2) и использует статистику

$$c = \frac{\max\limits_{1 \leqslant i \leqslant k} \omega_i}{\sum\limits_{i=1}^{k} \omega_i},$$

где $\omega_i = \max\limits_{1 \leqslant j \leqslant n} x_{ji} - \min\limits_{1 \leqslant j \leqslant n} x_{ji}$ — размах i-й выборки.

Если $c > c_\alpha(n,k)$, то нулевая гипотеза отклоняется. Критические значения $c_\alpha(n,k)$ приведены в табл. 135. Критерий не обладает высокой мощностью и рекомендуется к применению при $n \leqslant 10$.

Таблица 135

Критические значения $c_\alpha(n,k)$ критерия Блисса–Кохрана–Тьюки для доверительной вероятности $\alpha = 0{,}95$ [119]

k	n							
	3	4	5	6	7	8	9	10
3	0,667	0,601	0,563	0,539	0,521	0,507	0,498	0,489
4	0,538	0,479	0,446	0,425	0,410	0,398	0,389	0,382
5	0,451	0,398	0,369	0,351	0,338	0,328	0,320	0,314
6	0,389	0,342	0,316	0,300	0,288	0,280	0,273	0,267
7	0,342	0,300	0,278	0,263	0,353	0,245	0,239	0,234
8	0,305	0,267	0,248	0,234	0,225	0,218	0,213	0,208
9	0,276	0,241	0,224	0,211	0,203	0,197	0,192	0,188
10	0,253	0,220	0,204	0,193	0,185	0,179	0,184	0,172

Задача 208. *Проверить нулевую гипотезу равенства дисперсий критерием Блисса–Кохрана–Тьюки в условиях задачи* 205.

Имеем $\omega_1 = 7 - 3 = 4$; $\omega_2 = 15 - 2 = 13$; $\omega_3 = 28 - 9 = 19$; $\omega_4 = 16 - 4 = 12$.

Тогда $c = \dfrac{\max\limits_{1 \leqslant i \leqslant 4} \omega_i}{\sum\limits_{i=1}^{4} \omega_i} = \dfrac{19}{4 + 13 + 19 + 12} = 0{,}396.$

Из табл. 135 находим $c_{0,95}(5;4) = 0{,}446$.

Так как $c = 0{,}396 < c_{0,95}(5;4) = 0{,}446$, нулевая гипотеза не отклоняется.

4.1.1.4.5. Критерий Хартли

Критерий основан на статистике [416]

$$h = \frac{\max\limits_{1 \leqslant i \leqslant k} s_i^2}{\min\limits_{1 \leqslant i \leqslant k} s_i^2}.$$

При $h > h_\alpha(n,k)$ нулевая гипотеза равенства дисперсий отклоняется. Критические значения $h > h_\alpha(n,k)$ приведены в [25, 119, 140] и воспроизведены в табл. 136. Критерий не обладает высокой мощностью и рекомендуется к применению при $k \leqslant 10$.

Таблица 136

Критические значения $h_\alpha(n,k)$ критерия Хартли для доверительной вероятности $\alpha = 0{,}95$ [119]

n	k										
	2	3	4	5	6	7	8	9	10	11	12
3	39,0	87,5	142	202	266	333	403	475	550	626	704
4	15,4	27,8	39,2	50,7	60,0	72,9	83,5	93,9	104	114	124
5	9,60	15,5	20,6	25,2	29,5	33,5	37,5	41,1	44,6	48,0	51,4
6	7,15	10,8	13,7	16,3	18,7	20,8	22,9	24,7	26,5	28,2	29,9
7	5,82	8,38	10,4	12,1	13,7	15,0	16,3	17,5	18,6	19,7	20,7
8	4,99	6,94	8,44	9,70	10,8	11,8	12,7	13,5	14,3	15,1	15,8
9	4,43	6,00	7,18	8,12	9,03	9,78	10,5	11,1	11,7	12,2	12,7
10	4,03	5,34	6,31	7,11	7,80	8,41	8,95	9,45	9,91	10,3	10,7
13	3,28	4,16	4,79	5,30	5,72	6,09	6,42	6,72	7,00	7,25	7,48
16	2,86	3,54	4,01	4,37	4,68	4,95	5,19	5,40	5,59	5,77	5,93
21	2,46	2,95	3,29	3,54	3,76	3,94	4,10	4,24	4,37	4,49	4,59
31	2,07	2,40	2,61	2,78	2,91	3,02	3,12	3,21	3,29	3,36	3,39
61	1,67	1,85	1,96	2,04	2,11	2,17	2,22	2,26	2,30	2,33	2,36

Задача 209. *Проверить нулевую гипотезу равенства дисперсий критерием Хартли в условиях задачи* 205.

Имеем $\max s_i^2 = s_3^2 = 58{,}3$; $\min s_i^2 = s_1^2 = 2{,}5$; $h = \dfrac{58{,}3}{2{,}5} = 23{,}32$.

Из табл. 136 для $k = 4$ и $n = 5$ имеем $h_{0,95}(5;4) = 20{,}6$.

Так как $h = 23{,}32 > h_{0,95}(5;4) = 20{,}62$, нулевая гипотеза отклоняется (т.е. дисперсии не признаются равными).

4.1.1.4.6. Критерий Кэдуэлла–Лесли–Брауна

Кэдуэлл [147], Лесли и Браун [418] предложили аналог критерия Хартли (см. раздел 4.1.1.4.5), основанный на статистике отношения размахов

$$K = \frac{\max\limits_{1\leqslant i\leqslant k} \omega_i}{\min\limits_{1\leqslant i\leqslant k} \omega_i}.$$

Критические значения статистики $K_\alpha(n,k)$, превышение которых приводит к отклонению нулевой гипотезы, приведены в табл. 137. Критерий применим при $k \leqslant 10$, однако его мощность меньше мощности критерия Хартли.

Таблица 137

Критические значения $K_\alpha(n,k)$ критерия Кэдуэлла–Лесли–Брауна для доверительной вероятности $\alpha = 0{,}95$ [119]

n	k								
	2	3	4	5	6	7	8	9	10
3	6,28	9,32	11,90	14,20	16,30	18,20	20,00	21,70	23,30
4	3,96	5,31	6,32	7,20	7,95	8,63	9,24	9,76	10,30
5	3,15	4,02	4,63	5,10	5,53	5,93	6,26	6,55	6,80
6	2,74	3,37	3,82	4,16	4,47	4,71	4,93	5,14	5,32
7	2,49	2,99	3,34	3,61	3,85	4,04	4,22	4,37	4,51
8	2,32	2,75	3,04	3,27	3,46	3,62	3,75	3,88	3,99
9	2,20	2,58	2,83	3,03	3,19	3,32	3,44	3,55	3,64
10	2,11	2,45	2,68	2,84	2,99	3,11	3,21	3,31	3,39

Задача 210. *Проверить нулевую гипотезу равенства дисперсий критерием Кэдуэлла—Лесли—Брауна в условиях задачи* 205.

Имеем $\max\limits_{1\leqslant i\leqslant 4}\omega_i = \omega_3 = 28 - 9 = 19$; $\min\limits_{1\leqslant i\leqslant 4}\omega_i = \omega_1 = 7 - 3 = 4$.

Тогда $K = \dfrac{19}{4} = 4{,}75$. Из табл. 137 имеем $K_{0,95}(5;4) = 4{,}63$.

Так как $K = 4{,}75 > K_{0,95}(5;4) = 4{,}63$, нулевая гипотеза отклоняется.

4.1.1.4.7. Критерий Самиуддина

Критерий Бартлетта (см. раздел 4.1.1.4.1) очень чувствителен к отклонениям от нормальности. Поэтому не прекращаются попытки найти альтернативные ему критерии, отличающиеся повышенной устойчивостью к отклонениям от нормальности. Одна из таких попыток рассмотрена в работах [419, 420]. В них Самиуддином предложены статистики, основанные на использовании корня кубического из выборочной дисперсии, исходя из того факта, что отношение $\left(\dfrac{s_i^2}{\sigma_i^2}\right)^{\frac{1}{3}}$ распределено асимптотически нормально.

Пусть s_i^2 — выборочные оценки дисперсии с $f_i = n_i - 1$ степенями свободы. Обозначим

$$t_i \ln s_i^2; \quad \bar{t} = \frac{1}{f}\sum_{i=1}^{k} f_i t_i; \quad f = \sum_{i=1}^{k} f_i \bar{s}^2 = \frac{1}{f}\sum_{i=1}^{k} f_i s_i^2.$$

Тогда $B = f\left(\ln\bar{x}^2 - \bar{t}\right)$, есть статистика критерия Бартлетта, рассмотренная в разделе 4.1.1.4.1. Для записи статистик критериев введем обозначения

$$t_i = \left(s_i^2\right)^{\frac{1}{3}} b_i = \frac{2}{9f}; \quad T = \left\{\frac{1}{f}\sum_{i=1}^{k} f_i s_i^2\right\}^{\frac{1}{3}}; \quad z_i = \frac{f_i s_i^2}{\sum\limits_{i=1}^{k} f_i s_i^2}.$$

Тогда статистики модифицированных критериев Самиуддина имеют вид

$$W^* = \sum_{i=1}^{k}\frac{\left[t_i - (1 - b_i)\,T\right]^2}{b_i T^2}; \quad W^{**} = \frac{9}{2}\sum_{i=1}^{k} f_i\left[\left(\frac{f}{f_i}\right)^{\frac{1}{3}} z_i^{\frac{1}{3}} - 1\right]^2.$$

При $k \gg 1$ статистики W^* и W^{**} имеют распределение $\chi^2(k-1)$. Нулевая гипотеза равенства дисперсий отклоняется, если

$$W^*(W^{**}) > \chi^2_\alpha(k-1),$$

где $\chi^2_\alpha(k-1)$ — α-квантиль распределения χ^2 с $(k-1)$ степенями свободы (ее значения приведены в табл. 55).

Из предложенных критериев наибольшей мощностью обладает критерий W^{**}, но все они превышают по мощности критерий Бартлетта.

Задача 211. *Проверить нулевую гипотезу равенства дисперсий в условиях задачи* 205, *используя семейство критериев Самиуддина* ($\alpha = 0{,}95$).

Имеем

$$f_1 = \ldots = f_4 = n - 1 = 5 - 1 = 4; \quad s_1^2 = 2{,}5; \quad s_2^2 = 22{,}5; \quad s_3^2 = 58{,}3; \quad s_4^2 = 21{,}2;$$

$$t_1\left(s_1^2\right)^{\frac{1}{3}} = 1{,}3572; \quad t_2 = \left(s_2^2\right)^{\frac{1}{3}} = 2{,}8231; \quad t_3 = \left(s_3^2\right)^{\frac{1}{3}} = 3{,}8775; \quad t_4 = \left(s_4^2\right)^{\frac{1}{3}} = 2{,}7676;$$

$$b_1 = \ldots = b_4 = \sqrt{\frac{2}{9\cdot 4}} = 0{,}2357.$$

Далее

$$T = \left\{ \frac{1}{f} \cdot \sum_{i=1}^{4} f_i^2 \cdot s_i^2 \right\}^{\frac{1}{3}} = \left\{ \frac{4}{4 \cdot 5} \cdot 104{,}5 \right\}^{\frac{1}{3}} = 2{,}7545;$$

$$z_1 = \frac{f_1 \cdot s_1^2}{\sum\limits_{i=1}^{4} f_i \cdot s_i^2} = \frac{4 \cdot 2{,}5}{4 \cdot 104{,}5} = 0{,}0239; \quad z_2 = 0{,}2153; \quad z_3 = 0{,}5579; \quad z_4 = 0{,}2029.$$

Вычисляем статистики критериев:

$$W^* = \sum_{i=1}^{4} \frac{[t_i - (1 - b_i) \cdot T]^2}{b_i \cdot T} = \frac{1}{0,2357 \cdot 2,7545^2} \cdot \sum_{i=1}^{4} \left[t_i - (1 - 0,2357) \cdot 2,7545^2\right] = 2,6027;$$

$$W^{**} = \frac{9}{2} \cdot 4 \cdot \sum_{i=1}^{4} \left[\left(\frac{4 \cdot 5}{4} \right)^{\frac{1}{3}} \cdot z_i^{\frac{1}{3}} \right]^2 = 7,638.$$

Из табл. 55 для $f = k - 1 = 3$ и $\alpha = 0{,}95$ находим $\chi^2_{0,95}(3) = 7{,}81$.

Так как $W^* = 2{,}6027$ и $W^{**} = 7{,}638$ меньше, чем $\chi^2_{0,95}(3) = 7{,}81$, нулевая гипотеза не отклоняется.

4.1.2. Сравнение параметров экспоненциальных распределений

4.1.2.1. Сравнение двух параметров

Предположим, имеются две выборки случайных величин (например, наработка на отказ изделия) объемами n и m из экспоненциальных распределений (см. раздел 1.1.4)

$$x_1, x_2, \ldots, x_n \quad \text{и} \quad y_1, y_2, \ldots, y_m,$$

т. е. из распределений с плотностями

$$f(x) = \frac{1}{\nu_1} \exp\left(-\frac{x}{\nu_1}\right) \quad \text{и} \quad f(y) = \frac{1}{\nu_2} \exp\left(-\frac{y}{\nu_2}\right),$$

где ν_1 и ν_2 — параметры распределений (средние значения).

Иногда в практике (задачи анализа надежности объектов) используют параметр $\lambda = \dfrac{1}{\nu}$ — интенсивность отказов. В дальнейшем мы будем уточнять, о каком параметре идет речь.

4.1.2.1.1. Критерий Фишера

Статистика критерия имеет вид

$$F = \frac{m \sum\limits_{i=1}^{n} x_i}{n \sum\limits_{j=1}^{k} y_j}.$$

При справедливости нулевой гипотезы статистика F имеет распределение Фишера (см. раздел 1.1.10) с $f = 2n$ и $f = 2m$ степенями свободы. Если F_γ — γ-квантиль распределения Фишера, то с достоверностью α нулевая гипотеза отклоняется:

в пользу альтернативы $H_1 : \nu_1 \neq \nu_2$, если

$$F \geqslant F_{\frac{1+\alpha}{2}}(2n, 2m) \quad \text{или} \quad F \leqslant F_{\frac{1-\alpha}{2}}(2n, 2m)\,;$$

в пользу альтернативы $H_1' : \nu_1 > \nu_2$, если $F > F_\alpha(2n, 2m)$; в пользу альтернативы $H_1'' : \nu_1 < \nu_2$, если $F < F_{1-\alpha}(2n, 2m)$.

Задача 212. *Две партии приборов были испытаны на надежность. В результате были получены следующие значения моментов их отказов (в условных единицах):*

$$x_i\colon 12,\ 14,\ 16,\ 20,\ 30,\ 40,\ 60,\ 85\ (n = 8);$$
$$y_j\colon 22,\ 38,\ 44,\ 54,\ 68,\ 72\ (m = 6).$$

Необходимо проверить нулевую гипотезу равенства средних $H_0 : \nu_1 = \nu_2$ против альтернативы $H_1 : \nu_1 \neq \nu_2$ при доверительной вероятности $\alpha = 0{,}95$.

Имеем $\sum x_i = 277$; $\sum y_j = 298$; $F = \dfrac{6 \cdot 277}{8 \cdot 298} = 0{,}697$.

Так как в числителе должно быть большее значение, используем величину $F = \dfrac{n \cdot \sum y_j}{m \cdot \sum x_i} = 1{,}434$.

При необходимости найти критическое значение при $\alpha = 0{,}95$: $F_{\frac{1+0{,}95}{2}}(2m; 2n) = F_{0{,}975}(12; 16)$.

По таблицам F-распределения (подробнее см. раздел 1.1.10) находим $F_{0{,}975}(12{,}16) = 2{,}90$. Так как $F = 1{,}434 < F_{0{,}975}(12{,}16) = 2{,}90$, нулевая гипотеза не отклоняется.

4.1.2.1.2. Критерий Фишера при сравнении интенсивностей отказов (λ)

Рассмотрим сравнение интенсивностей отказов λ в двух экспоненциально распределенных выборках. Это наиболее распространенный показатель надежности, используемый при оценке технических объектов. Обычной ситуацией, в которой получается оценка интенсивности отказов, является испытание объектов, в процессе которых фиксируется количество отказов за определенный временной промежуток. Именно в такой интерпретации изложим применение критерия Фишера.

Предположим, были зафиксированы за время испытаний t_1 первого объекта r_1 отказов, а у второго объекта r_2 отказов за время t_2.

Для проверки нулевой гипотезы $H_0 : \lambda_1 = \lambda_2$ используется статистика $R = \dfrac{r_1 t_2}{r_2 t_1} = \dfrac{\lambda_2}{\lambda_1}$, имеющая при справедливости нулевой гипотезы F-распределение с $f_1 = 2r_2$ и $f_2 = 2r_1$ степенями свободы. При справедливости нулевой гипотезы (т. е., когда $\dfrac{\lambda_2}{\lambda_1} \approx 1$) должно выполняться неравенство

$$R = \frac{r_1 t_2}{r_2 t_1} < F_{\frac{1+\alpha}{2}}(2r_2, 2r_1)\,.$$

Задача 213. *При испытании первого изделия были получены 12 отказов за 210 ч испытаний. У второго изделия были получены 5 отказов за 100 ч испытаний. Проверить гипотезу равенства интенсивностей отказов у изделий при доверительной вероятности $\alpha = 0{,}95$.*

Имеем $r_1 = 5$; $t_1 = 100$; $r_2 = 12$; $t_2 = 210$; $R = \dfrac{12 \cdot 100}{5 \cdot 210} = 1{,}143$.

Из таблиц находим $F_{0{,}975}(10{,}24) = 2{,}64$.

Так как $R = 1{,}143 < F_{0{,}975}(10{,}24) = 2{,}64$, нулевая гипотеза не отклоняется.

4.1.2.1.3. Двухвыборочный пуассоновский критерий

Известно (см. раздел 1.1.4), что экспоненциальное распределение порождается пуассоновским потоком с интенсивностью λ.

Предположим, наблюдаются r_1 отказов одного изделия за время t_1 и r_2 отказов второго изделия за время t_2. В нашем случае r_1 и r_2 — случайные величины, подчиненные закону Пуассона с параметрами $\gamma_1 = t_1\lambda_1$ и $\gamma_2 = t_2\lambda_2$.

В [421] предложен критерий проверки нулевой гипотезы $H_0\colon \lambda_1 = \lambda_2$ против альтернативы $H_1\colon \lambda_1 < \lambda_2$, основанный на статистике $z = \left(r_2 - \frac{t_2}{t_1}r_1\right)\left[\frac{t_2}{t_1}(r_1+1)\right]^{-\frac{1}{2}}$.

Нулевая гипотеза отклоняется, если $z > u_\alpha$, где u_α — α-квантиль стандартного нормального распределения, α — уровень достоверности.

Более точный критерий предложен в [422]:

$$z^* = \frac{2\cdot\left\{\left(r_2+\frac{3}{8}\right)^{\frac{1}{2}} - \left[\frac{t_2}{t_1}\cdot\left(r_1+\frac{3}{8}\right)\right]^{\frac{1}{2}}\right\}}{\left(1+\frac{t_2}{t_1}\right)^{\frac{1}{2}}}.$$

При $z^* > u_\alpha$ он также отклоняет нулевую гипотезу.

Задача 214. *В условиях задачи* 213 *проверить нулевую гипотезу равенства интенсивностей отказов двухвыборочным пуассоновским критерием при уровне достоверности* $\alpha = 0{,}95$.

Имеем $r_1 = 5$; $t_1 = 100$; $r_2 = 12$; $t_2 = 210$; $z = \dfrac{12 - \frac{210}{100}\cdot 5}{\sqrt{\frac{210}{100}\cdot(5+1)}} = 0{,}422$.

Из табл. 1 находим $u_{0,95} = 1{,}64$. Так как $z = 0{,}422 < u_{0,95} = 1{,}64$, нулевая гипотеза не отклоняется.

По аналогии находим

$$z^* = \frac{2\cdot\left\{\left(12+\frac{3}{8}\right)^{\frac{1}{2}} - \left[\frac{210}{100}\cdot\left(5+\frac{3}{8}\right)\right]^{\frac{1}{2}}\right\}}{\left(1+\frac{210}{100}\right)^{\frac{1}{2}}} = 0{,}179,$$

что также не отклоняет нулевую гипотезу.

4.1.2.1.4. Сравнение значения параметра с заданным

В практике испытаний на надежность технических объектов особый интерес представляет проверка гипотез $H_0\colon \nu \leqslant \nu_0$ или $H_0\colon \nu > \nu_0$, где ν_0 — некоторое заданное значение наработки на отказ, являющееся, как правило, нормируемым количественным показателем надежности.

Рассмотрим методы проверки гипотез о значении интенсивности отказов изделий $\left(\lambda = \frac{1}{\nu}\right)$ применительно к различным планам испытаний на надежность (подробнее о планах испытаний на надежность см. в разделе 2.2).

План $[\boldsymbol{N}, \boldsymbol{B}, \boldsymbol{T}]$

Статистикой критерия проверки нуль-гипотезы является число отказов r за время испытаний T. Если $r \leqslant c$, где c — приемочное число, удовлетворяющее

неравенству

$$K_\alpha(c) \geqslant N\lambda_0 T,$$

где $K_\alpha(c)$ — коэффициент, значения которого для различных α и c приведены в табл. 138; λ_0 — гипотетическое значение интенсивности отказов; α — доверительная вероятность, то гипотеза $H_0\colon \lambda > \lambda_0$ отклоняется в пользу альтернативы $H_1 : \lambda \leqslant \lambda_0$ с достоверностью α.

Таблица 138

Значения коэффициентов $K_\alpha(c)$ [140]

c	Доверительная вероятность α			
	0,99	0,95	0,93	0,90
0	0,01005	0,05129	0,07257	0,10536
1	0,14855	0,35536	0,43081	0,53181
2	0,43604	0,81769	0,94223	1,10206
3	0,82325	1,36632	1,53414	1,74477
4	1,27911	1,97015	2,17670	2,43259
5	1,78528	2,61301	2,85488	3,15190
6	2,33021	3,28532	3,55984	3,89477
7	2,90611	3,98082	4,28584	4,65612
8	3,50746	4,69523	5,02895	5,43247
9	4,13020	5,42541	5,78633	6,22130
10	4,77125	6,16901	6,55583	7,02075
11	5,42818	6,92421	7,33581	7,82934
12	6,09907	7,68958	8,12496	8,64594
13	6,78235	8,46394	8,92222	9,46962
14	7,47673	9,24633	9,72672	10,29962
15	8,18111	10,03596	10,53773	11,13530
16	8,89457	10,83114	11,35465	11,97613

План $[N, B, r]$

Статистикой критерия является суммарная наработка изделий до момента наступления r-го отказа $T_r = Nr$, критические значения которой находятся по формуле

$$T_r(\alpha) = \frac{1}{\lambda_0} K_\alpha(r-1)\,,$$

где $K_\alpha(r-1)$ — коэффициент из табл. 138 при $c = r - 1$.

Гипотеза $H_0\colon \lambda < \lambda_0$ отклоняется в пользу альтернативы $H_1\colon \lambda \geqslant \lambda_0$, если $T_r < T_r(\alpha)$.

План $[N, Б, T]$

Статистикой критерия проверки нулевой гипотезы является число отказов r, полученное за время испытаний T.

Если $r > c$, где c — наименьшее целочисленное решение неравенства

$$\sum_{i=c+1}^{N} C_N^i \left(1 - e^{-\lambda_0 T}\right)^i e^{-\lambda_0 T(N-i)} \leqslant \alpha,$$

то с вероятностью α гипотеза $H_0\colon \lambda \leqslant \lambda_0$ отклоняется в пользу альтернативы $H_1 : \lambda > \lambda_0$.

План $[N, Б, r]$

Статистикой критерия для проверки нулевой гипотезы является суммарная наработка

$$T_r = \sum_{i=1}^{r} t_i + (N - r)\, t_r,$$

критические значения которой равны

$$T_r(\alpha) = \frac{1}{\lambda_0} K_\alpha(r-1)\,.$$

Если $T_r < T_r(\alpha)$, то гипотеза $H_0\colon \lambda < \lambda_0$ отклоняется в пользу альтернативы $H_1\colon \lambda \geqslant \lambda_0$.

Задача 215. *Партия электронных приборов была испытана на надежность в соответствии с планом испытаний $[N, B, T]$. При этом испытаниям подвергались $N = 100$ ламп в течение $T = 200$ ч с заменой отказавших ламп. При испытаниях были зафиксированы $r = 5$ отказов. Необходимо с достоверностью $\alpha = 0{,}99$ проверить гипотезу о том, что интенсивность отказов ламп не превышает величины $\lambda_0 = 10^{-4}$ ч$^{-1}$.*

Вычисляем $N \cdot \lambda_0 \cdot T = 100 \cdot 10^{-4} \cdot 200 = 2$.

Из табл. 138 для $\alpha = 0{,}99$ видим, что $K_{0,99}(c) \geqslant 2$ при $c = 6$.

Так как $r = 5 < c = 6$, гипотеза $H_0\colon \lambda \leqslant \lambda_0$ не отклоняется.

Находим приемочное число c из условия $K_\alpha(c) \geqslant N \cdot \lambda_0 \cdot T = 2$.

Задача 216. *На надежность была испытана партия электронных ламп в соответствии с планом испытаний $[N, B, r]$. Испытаниям подвергались $N = 100$ ламп с заменой отказавших ламп. Испытания проводились до появления $r = 5$ отказов. При этом пятый отказ наступил при $t_5 = 200$ ч. Необходимо с достоверностью $\alpha = 0{,}95$ проверить нулевую гипотезу о том, что интенсивность отказов ламп не превосходит допустимой величины $\lambda_0 = 1{,}3 \cdot 10^{-4}$ ч$^{-1}$.*

Из табл. 138 для $\alpha = 0{,}95$ и $r = 5$ $(c = r - 1 = 4)$ находим $K_{0,95}(4) = 1{,}97015$ и

$$T_5(0{,}95) = \frac{K_{0,95}(4)}{\lambda_0} = \frac{1{,}97015}{1{,}3 \cdot 10^{-4}} = 15155 \text{ ч}.$$

В нашем случае $T_r = N \cdot t_r = 100 \cdot 200 = 20000$ ч.

Так как $T_5 = 20000 > T_5(0{,}95) = 15155$, нулевая гипотеза $H_0\colon \lambda < \lambda_0$ не отклоняется, т. е. с достоверностью 0,95 можно утверждать, что интенсивность отказов ламп не превосходит $1{,}3 \cdot 10^{-4}$ ч$^{-1}$.

Задача 217. *Проверить нулевую гипотезу в условиях задачи 216, если испытания проводились без замены отказавших приборов, т. е. когда вместо плана $[N, B, r]$ был реализован план $[N, Б, r]$. При этом были зафиксированы моменты наступления отказов ламп*

$$t_1 = 50, \quad t_2 = 60, \quad t_3 = 100, \quad t_4 = 180, \quad t_5 = 200 \text{ ч}.$$

Имеем суммарную наработку до 5-го отказа

$$T_5 = \sum_{i=1}^{5} t_i + (100 - 5)\, t_5 = 50 + 60 + 100 + 180 + 200 + 95 \cdot 200 = 19590 \text{ ч}.$$

Критическое значение равно

$$T_5(0{,}95) = \frac{K_{0,95}(4)}{1{,}3 \cdot 10^{-4}} = \frac{1{,}97015}{1{,}3 \cdot 10^{-4}} = 15155.$$

Так как $T_r = 19590 > T_5(0,95) = 15155$, нулевая гипотеза не отклоняется.

Задача 218. *Партия изделий $N=5$ была испытана на надежность в течение $T=100$ ч без замены отказавших приборов (т. е. в соответствии с планом $[N, Б, T]$). При этом были получены $r=2$ отказа. Необходимо проверить гипотезу $H_0\colon \lambda<\lambda_0$ при доверительной вероятности $\alpha=0{,}95$ и $\lambda_0=6{,}93\cdot 10^{-3}$ ч$^{-1}$.*

Находим подбором наименьшее целочисленное решение неравенства

$$\sum_{i=c+1}^{N} C_N^i\cdot\left(1-e^{-\lambda_0\cdot T}\right)^i\cdot e^{-\lambda_0\cdot T\cdot(N-i)}\leqslant\alpha.$$

В нашем случае

$$\sum_{i=c+1}^{5} C_5^i\cdot\left(1-e^{-6{,}93\cdot 10^{-3}\cdot 100}\right)\cdot e^{-6{,}93\cdot 100\cdot(5-i)}\leqslant 0,95;$$

$$\sum_{i=c+1}^{5} C_5^i\cdot 0,5^i\cdot 0,5^{N-i}=\sum_{i=c+1}^{5} C_5^i\cdot 0,5^N=0,5\cdot 5\cdot\sum_{i=c+1}^{5} C_5^i\leqslant 0,95;$$

$$\sum_{i=c+1}^{5} C_5^i\leqslant\frac{0,95}{0,5^5}=30,4.$$

Имеем при $c=0$: $\sum\limits_{i=1}^{5} C_5^i=5+10+10+5+1=31$; при $c=1$: $\sum\limits_{i=2}^{5} C_5^i=26<30{,}4$. Следовательно, $c=1$, а так как $r=2>c=1$, нулевая гипотеза отклоняется.

Задача 219. *Партия $N=15$ изделий была испытана на надежность без замены отказавших ламп до получения $r=5$ отказов. Моменты появления отказов равны (ч): $t_1=100$, $t_2=150$, $t_3=300$, $t_4=400$ и $t_5=510$ (т. е. имеет место план испытаний $[N, Б, r]$. Необходимо проверить гипотезу о том, что $\lambda<\lambda_0=1{,}8\cdot 10^{-4}$ ч$^{-1}$ при $\alpha=0{,}99$.*

Имеем $T_r=\sum\limits_{i=1}^{5} t_i+(15-5)\cdot 510=6560$.

Из табл. 138 находим $K_{0{,}99}(5-1)=K_{0{,}99}(4)=1{,}279$.

Критическое значение равно $T_5(0{,}99)=\dfrac{1{,}279}{1{,}8\cdot 10^{-4}}=7105{,}5$.

Так как $T_5=6560<T_5(0,99)=7105$, нулевая гипотеза отклоняется и можно утверждать, что интенсивность отказов изделия превышает $1{,}8\cdot 10^{-4}$ ч$^{-1}$.

4.1.2.2. Сравнение нескольких ($k\geqslant 2$) параметров

4.1.2.2.1. Критерий Дэвида

Предположим, что k выборок одинакового объема n испытаны в течение фиксированного времени t. При этом зафиксированы r_i отказов в каждой i-й выборке. Полученные значения r_i упорядочим по убыванию $r_1\geqslant r_2\geqslant\ldots\geqslant r_k$.

Дэвид [421] предложил для проверки гипотезы $H_0\colon \lambda_i=\lambda_{i+1}$ $(i=1,2,\ldots,k-1)$ критерий, основанный на статистике

$$D_i=2\left\{\left(r_i+\frac{1}{2}\right)^{\frac{1}{2}}-\left(r_{i+1}+\frac{1}{2}\right)^{\frac{1}{2}}\right\}.$$

При $D_i>D_\alpha$, где D_α — критическое значение статистики, приведенное в табл. 139, гипотеза H_0 отклоняется в пользу альтернативы $H_1\colon \lambda_1=\lambda_2=\ldots=\lambda_i\neq\lambda_{i+1}=\ldots=\lambda_k$.

Если все $D_i<D_\alpha$, то гипотеза H_0 принимается на уровне достоверности α.

Таблица 139

Критические значения D_α критерия Дэвида
(α — доверительная вероятность) [421]

k	2	3	4	5	6	7	8	9	10	11	12
$D_{0,95}$	2,77	2,48	2,28	2,13	2,02	1,93	1,86	1,81	1,76	1,72	1,69

Задача 220. *При испытаниях пяти выборок изделий объемом $n = 10$ каждая, в течение одинакового промежутка времени были получены следующие количества отказов в выборках: $r_1 = 11$, $r_2 = 9$, $r_3 = 5$, $r_4 = 4$ и $r_5 = 2$. Необходимо проверить гипотезу равенства интенсивностей отказов в выборках при достоверности $\alpha = 0{,}95$.*

Имеем ряд значений

$$D_1 = 2\cdot\sqrt{r_1+0{,}5} - 2\cdot\sqrt{r_2+0{,}5} = 2\cdot\sqrt{11{,}5} - 2\cdot\sqrt{9{,}5} = 0{,}618;$$
$$D_2 = 1{,}474; \quad D_3 = 0{,}448; \quad D_4 = 1{,}080.$$

Из табл. 139 для $k = 5$ имеем $D_{0,95} = 2{,}13$. Так как все $D_i < D_{0,95}$, то нулевая гипотеза равенства интенсивностей отказов не отклоняется.

4.1.2.2.2. Критерий максимального правдоподобия

Статистика критерия имеет вид

$$H = 2\left\{\sum_{i=1}^{k} r_i \ln r_i - \sum_{i=1}^{k} r_i \ln \frac{1}{k}\sum_{i=1}^{k} r_i\right\}.$$

При справедливости нулевой гипотезы H-статистика распределена как χ^2 с $f = k - 1$ степенями свободы. Поэтому нулевая гипотеза равенства k интенсивностей отказов не отклоняется при доверительной вероятности α, если $H < \chi^2_\alpha(f)$, где значения $\chi^2_\alpha(f)$ приведены в табл. 55.

Если в качестве наблюдаемой случайной величины рассматриваются моменты наступления фиксированного числа отказов $r(t_r)$, то H-статистика используется в форме [422]

$$H^* = 2\left\{ r\sum_{i=1}^{k} \ln\frac{r}{t_{ri}} - k\ln\frac{kr}{\sum_{i=1}^{k} t_{ri}} \right\},$$

где t_{ri} — момент наступления r-го отказа в i-й выборке.

Задача 221. *При испытаниях пяти выборок изделий ($k = 5$) были получены 4 отказа в каждой выборке ($r = 4$). При этом моменты наступления последних по времени отказов в выборках равнялись*

$$t_1 = 120, \quad t_2 = 210, \quad t_3 = 340, \quad t_4 = 510, \quad t_5 = 520.$$

Необходимо проверить гипотезу равенства интенсивностей отказов в выборках критерием максимального правдоподобия при доверительной вероятности $\alpha = 0{,}95$.

Имеем

$$5H^* = 2\cdot\left\{ 4\cdot\ln\frac{4}{t_{ri}} - 5\cdot 4\cdot\ln\frac{5\cdot 4}{\sum_{i=1}^{5} t_{ri}} \right\} = 5{,}543.$$

Из табл. 5 для $\alpha = 0{,}95$ находим $\chi^2_{0,95}(f = 4) = 11{,}14$.

Так как $H^* = 5{,}54 < \chi^2_{0,95}(4) = 11{,}14$, нулевая гипотеза не отклоняется.

Задача 222. *Предположим, что в задаче 221 наблюдались не моменты времени до достижения фиксированного количества отказов в выборках, а количества отказов за фиксированное время испытаний. Пусть за одно и тоже время испытаний t в выборках наблюдались следующие количества отказов: $r_1 = 5$, $r_2 = 4$, $r_3 = 2$, $r_4 = 6$, $r_5 = 1$. Необходимо проверить гипотезу равенства интенсивностей отказов в выборках критерием максимального правдоподобия.*

Имеем $k = 5$; $\sum\limits_{i=1}^{5} r_i = 5 + 4 + 2 + 6 + 1 = 18$.

Тогда
$$H = 2 \cdot \left\{ \sum_{i=1}^{5} r_i \cdot \ln r_i - \sum_{i=1}^{5} r_i \cdot \ln\left(\frac{18}{5}\right) \right\} = 5{,}346.$$

Так как $H = 5{,}346 < \chi^2_{0,95} = 11{,}14$, нулевая гипотеза не отклоняется.

4.1.2.2.3. Критерий отношения правдоподобия (критерий Нагарсенкера)

Предположим, имеются k выборок объема n каждая, содержащие экспоненциально распределенные случайные величины. В [423] для проверки нулевой гипотезы $H_0\colon \nu_1 = \nu_2 = \ldots = \nu_k$ равенства средних значений во всех выборках рассмотрен критерий, основанный на статистике

$$L = \prod_{j=1}^{k} \frac{\bar{x}_j}{\bar{x}}, \quad \text{где} \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}; \quad \bar{x} = \frac{1}{nk} \sum_{i=1}^{k} x_{ij}.$$

Если $L \leqslant L_\alpha(n,k)$, то нулевая гипотеза отклоняется с вероятностью α. Критические значения $L_\alpha(n,k)$ для $\alpha = 0{,}95$ приведены в табл. 140.

Таблица 140

Критические значения $L_\alpha(n,k)$ критерия отношения правдоподобия для доверительной вероятности $\alpha = 0{,}95$ [423]

n	k				n	k			
	3	4	5	6		3	4	5	6
4	0,454				10	0,736	0,671	0,616	0,569
5	0,535	0,443			20	0,859	0,821	0,787	0,756
6	0,596	0,510	0,442		40	0,927	0,906	0,888	0,870
7	0,643	0,563	0,498	0,444	80	0,963	0,952	0,942	0,933
8	0,681	0,606	0,545	0,492	100	0,970	0,962	0,954	0,946
9	0,711	0,641	0,583	0,534					

Задача 223. *Имеются четыре ($k = 4$) выборки из экспоненциального распределения:*

x_{i1}: 2, 4, 8, 12, 18; x_{i2}: 5, 7, 11, 19, 21;
x_{i3}: 11, 17, 21, 22, 29; x_{i4}: 1, 5, 9, 13, 18.

Необходимо проверить гипотезу равенства параметров экспоненциальных распределений в выборках критерием Нагарсенкера при доверительной вероятности $\alpha = 0{,}95$.

Находим $x_1 = 8{,}8$; $\bar{x}_2 = 12{,}6$; $\bar{x}_3 = 20$; $\bar{x}_4 = 9{,}2$; $\bar{x} = \dfrac{1}{4 \cdot 5} \cdot \sum\limits_{j=1}^{4} \sum\limits_{i=1}^{5} x_{ij} = 12{,}65$.

Далее имеем $L = \prod\limits_{j=1}^{5} \dfrac{\bar{x}_j}{\bar{x}} = \dfrac{8{,}8}{12{,}65} \cdot \dfrac{12{,}6}{12{,}65} \cdot \dfrac{20}{12{,}65} \cdot \dfrac{9{,}2}{12{,}65} = 0{,}797$.

Из табл. 140 находим $L_{0,95}(5{,}4) = 0{,}443$.

Так как $L = 0{,}707 > L_{0,95}(5{,}4) = 0{,}443$, нулевая гипотеза не отклоняется.

4.1.2.2.4. Критерий Чена для двухпараметрических экспоненциальных распределений

Рассмотрим k независимых экспоненциально распределенных случайных величин с плотностью распределения вероятностей

$$f(x) = \frac{1}{\nu} \exp\left\{-\frac{(x-\beta_i)}{\nu}\right\}.$$

Параметр положения β_i в практике интерпретируется обычно как гарантийная наработка, в пределах которой отказы изделия не допускаются.

Чен [424] предложил критерий проверки гипотезы $H_0\colon \beta_i = \text{const}$, утверждающей, что гарантийная наработка у всех испытываемых партий изделий постоянна.

Процедура построения статистики критерия Чена включает в себя отбор из каждой выборки (объема n) наименьшего наблюдения $y_j = \min\limits_{1\leqslant i\leqslant n} x_{ij}$ и определение размаха совокупности значений y_j

$$\omega_y = \max_{1\leqslant j\leqslant k} y_j - \min_{1\leqslant j\leqslant k} y_j.$$

Статистика критерия Чена имеет вид

$$c = kn(n-12)\,\omega_y\left\{\sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} = y_j)\right\}^{-1}.$$

Если $c > c_\alpha(k,n)$, то нулевая гипотеза отклоняется на уровне значимости α. Критические значения статистики Чена $c_\alpha(k,n)$ приведены в табл. 141.

Таблица 141

Критические значения $c_\alpha(k,n)$ статистики Чена [424]

n	Уровень значимости $\alpha = 0{,}1$						n	Уровень значимости $\alpha = 0{,}05$					
	k							k					
	3	4	5	10	20	30		3	4	5	10	20	30
3	3,77	4,09	4,31	4,92	5,50	5,86	3	5,00	5,23	5,29	5,85	6,34	6,66
4	3,48	3,83	4,08	4,76	5,40	5,78	4	4,50	4,81	5,02	5,61	6,20	6,55
5	3,34	3,71	3,97	4,68	5,35	5,74	5	4,28	4,62	4,85	5,50	6,13	6,50
6	3,26	3,64	3,90	4,64	5,32	5,71	6	4,15	4,50	4,75	5,43	6,08	6,47
8	3,18	3,56	3,83	4,58	5,28	5,68	8	4,01	4,38	4,63	5,36	6,04	6,43
10	3,13	3,52	3,79	4,55	5,27	5,67	10	3,93	4,31	4,57	5,31	6,01	6,41
16	3,06	3,46	3,73	4,51	5,24	5,65	16	3,83	4,21	4,49	5,26	5,97	6,38
30	3,02	3,41	3,69	4,48	5,22	5,64	30	3,75	4,15	4,43	5,21	5,94	6,36
60	2,99	3,39	3,67	4,47	5,21	5,63	60	3,71	4,11	4,39	5,19	5,93	6,35

Задача 224. *В условиях задачи 223 проверить критерием Чена гипотезу равенства параметров положения (β_i) на уровне значимости $\alpha = 0{,}05$ (напоминаем — это означает, что мы будем принимать значимыми, а не случайными, значения статистики Чена, вероятность появления которых не превышает 0,05).*

Имеем $y_1 = \min\limits_{1\leqslant i\leqslant 5} x_{i1} = 2$; $y_2 = 5$; $y_3 = 11$; $y_4 = 1$; $\omega_y = 11 - 1 = 10$.

Далее

$$\sum_{j=1}^{4}\sum_{i=1=}^{5}(x_{ij}-y_j)=(2-2)+(8-2)+\ldots+(13-1)+(18-1)=158;$$

$$c=\frac{4\cdot 5\cdot(5-1)\cdot 10}{158}=5{,}063.$$

Из табл. 141 при $\alpha = 0{,}05$ находим $c_{0,05}(4{,}5) = 4{,}62$.

Так как $c = 5{,}063 > c_{0,05}(4{,}5) = 4{,}62$, нулевая гипотеза отклоняется и гарантийные наработки у сравниваемых совокупностей следует признать отличающимися.

4.1.2.2.5. Комбинированный критерий Сингха

Обычно мы рассматриваем параметр положения β (гарантийный период, когда отказы не допускаются) и параметр масштаба ν (средняя наработка, измеряемая от параметра положения — гарантийного срока).

Критерий Сингха [437] предполагает проверку гипотезы

$$H_0\colon \beta_1=\ldots=\beta_k=\beta;\quad \nu_1=\ldots=\nu_k=\nu$$

против общей альтернативы для совокупностей, подчиняющихся экспоненциальному закону распределения вероятностей,

$$f(x_i)=\frac{1}{\nu_i}\exp\left\{-\frac{x_i-\beta_i}{\nu_i}\right\},\quad i=1,\ldots,k;\quad \nu_i>0.$$

Различные варианты критериев проверки такого рода гипотез рассмотрены в [425–431]. Подробно все эти критерии рассмотрены в [432], там же предложен один из наиболее мощных критериев для множественного сравнения параметров экспоненциальных распределений.

Пусть $x_{1i}\leqslant x_{2j}\leqslant\ldots\leqslant x_{kj}$ — порядковые наблюдения j-й выборки ($j=1,2,\ldots,k$). Предположим, что мы располагаем только первыми (наименьшими) r_j наблюдениями из каждой j-й выборки.

Оценками максимального правдоподобия для параметров ν_j и β_j будут (см. раздел 2.2)

$$b_j=x_{1j};\quad \hat\nu_j=\frac{1}{r_j}\left[\sum_{i=1}^{r_j}(x_{ij}-x_{1j})+(n_j-r_j)\left(x_{r_j,j}-x_{1j}\right)\right].$$

Введем обозначения

$$s_j=\sum_{i=1}^{r_j}(x_{ij}-x_{1j})+(n_j-r_j)\bigl(x_{r_j,j}-x_{1j}\bigr);\quad S=\sum_{j=1}^{k}s_j;\quad s_j^*=\frac{s_j}{r_j-1};$$

$$S^{**}=\frac{S}{R-k};\quad S^*=\sum_{j=1}^{k}\left[\sum_{i=1}^{r_j}(x_{ij}-x_m)+(n_j-r_j)\bigl(x_{r_j,j}-x_m\bigr)\right],$$

где $x_m=\min(x_{11},x_{12},x_{13},\ldots,x_{1k})$; $R=\sum_{j=1}^{k}r_j$.

Известно [431], что случайные величины $\frac{2s_j^*}{\nu}$, $\frac{2S^{**}}{\nu}$ и $\frac{2S^*}{\nu}$ распределены как χ^2 соответственно с $f=2(r_j-1)$; $2(r-k)$ и $2\,(k-1)$ степенями свободы.

Введя обозначение

$$V = \sum_{j=2}^{k} \left(\sum_{l=j}^{k} n_l \right) (x_{1j} - x_{1,j-1}),$$

запишем статистику

$$U = \frac{(R-1)\,V}{(k-1)\,S^*},$$

которая распределена как статистика Фишера с $f_1 = 2(k-1)$ и $f_2 = 2(R-1)$ степенями свободы.

Если $U > F_\alpha[2(k-1); 2(R-1)]$, то с достоверностью α нулевая гипотеза $H_0\colon \nu_1 = \ldots = \nu_k = \nu$ (ν и β — не определены) отклоняется в пользу альтернативы, утверждающей, что при равных ν по крайней мере две из k выборок различаются параметрами положения β (т. е. гарантийным сроком).

Для нулевой гипотезы $H_0'\colon \nu_1 = \ldots = \nu_k = \nu$ (β не определено) против альтернативы, утверждающей, что по крайней мере в двух выборках отличаются параметры ν, может быть использована модификация критерия Бартлетта (см. раздел 4.1.1.4.1), статистика которого применительно к проверке нашей гипотезы имеет вид

$$Q = \frac{1}{c} \left[f \ln S^{**} - \sum_{j=1}^{k} f_i \ln s_j^* \right],$$

где

$$f = R - r; \quad f_j = r_j - 1; \quad c = 1 + \frac{1}{3(k-1)} \left(\sum_{j=1}^{k} \frac{1}{f_j} - \frac{1}{f} \right).$$

При $Q \geqslant \chi^2_\alpha(k-1)$ гипотеза H_0' отклоняется в пользу альтернативы ($\chi^2_\alpha(k-1)$ для разных α см. в табл. 55).

Совместной проверкой критериев U и Q осуществляется комплексная проверка гипотез для трехпараметрического экспоненциального распределения.

Задача 225. *В результате испытаний на надежность 4 выборок изделий получены следующие результаты:*

— *1-я выборка* ($j = 1$) — *из* $n_1 = 24$ *испытанных приборов известны* $r_1 = 8$ *первых наблюдений:*

$$x_{11} = 1, \quad x_{21} = 4, \quad x_{31} = 11, \quad x_{41} = 16, \quad x_{51} = 18, \quad x_{61} = 24, \quad x_{71} = 31, \quad x_{81} = 39;$$

— *2-я выборка* ($j = 2$) — *из* $n_2 = 30$ *испытанных приборов известны данные по* $r_2 = 5$ *первым наблюдениям*

$$x_{12} = 3, \quad x_{22} = 5, \quad x_{23} = 13, \quad x_{24} = 14, \quad x_{25} = 22;$$

— *3-я выборка* ($j = 3$) — *из* $n_3 = 10$ *испытанных приборов известны данные для первых трех наблюдений* ($r_3 = 3$):

$$x_{13} = 10, \quad x_{23} = 14, \quad x_{33} = 20;$$

— *4-я выборка* ($j = 4$) — *из* $n_4 = 18$ *испытанных приборов известны* $r_4 = 6$ *первых результатов:*

$$x_{14} = 20, \quad x_{24} = 25, \quad x_{34} = 31, \quad x_{44} = 42, \quad x_{54} = 44, \quad x_{64} = 49.$$

Необходимо проверить гипотезу о равенстве параметров положения β_j *и масштаба* ν_j *во всех выборках при доверительной вероятности* $\alpha = 0{,}95$.

Находим

$$s_1 = \sum_{i=1}^{8} (x_{i1} - x_{11}) + (24 - 8) \cdot (x_{81} - x_{11}) = 744;$$

$$s_2 = \sum_{i=1}^{5} (x_{i2} - x_{12}) + (30 - 5) \cdot (x_{52} - x_{12}) = 517;$$

$$s_3 = \sum_{i=1}^{3} (x_{i3} - x_{13}) + (10 - 3) \cdot (x_{33} - x_{13}) = 84;$$

$$s_4 = \sum_{i=1}^{6} (x_{i4} - x_{14}) + (18 - 6) \cdot (x_{64} - x_{14}) = 439.$$

Далее вычисляем:

$$S = \sum_{j=1}^{k} s_j = 744 + 517 + 84 + 439 = 1784; \quad s_1^* = \frac{s_1}{8-1} = \frac{744}{7} = 106{,}2854;$$

$$s_2^* = \frac{s_2}{4} = 129{,}25; \quad s_3^* = \frac{s_3}{2} = \frac{84}{2} = 42; \quad s_4^* = \frac{s_4}{5} = \frac{439}{5} = 87{,}8;$$

$$S^* = \sum_{j=1}^{4} \left[\sum_{i=1}^{r_j} (x_{1,j} - 1) + (n_j - r_j) \cdot \left(x_{r_j,j} - 1 \right) \right] = 2276; \quad S^{**} = \frac{S}{\sum\limits_{j=1}^{k} r_j - k} = 99{,}111;$$

$$V = \sum_{j=2}^{k} \left(\sum_{i=l}^{k} n_l \right) \cdot (x_{ij} - x_{1,j} - 1) = 492; \quad U = \frac{(22-1) \cdot 492}{(2-1) \cdot 2276} = 1{,}513.$$

Находим из таблиц F-распределения $F_{0,95}(2 \cdot 3; 2 \cdot 21) = F_{0,95}(6, 42) = 2{,}33$.
Так как $U = 1{,}513 < F_{0,95}(6{,}42) = 2{,}33$, нулевая гипотеза H_0 не отклоняется.
Теперь для проверки нулевой гипотезы H_0' вычислим

$$f = R - k = 22 - 4 = 18; \quad f_1 = r_1 - 1 = 7; \quad f_2 = r_2 - 1 = 4;$$
$$f_3 = r_3 - 1; \quad f_4 = r_4 - 1 = 5;$$

$$c = 1 + \frac{1}{3 \cdot (4-1)} \cdot \left(\sum_{i=1}^{4} \frac{1}{f_i} - \frac{1}{18} \right) = 1 + \frac{1}{9} \cdot \left(\frac{1}{7} + \frac{1}{4} + \frac{1}{2} + \frac{1}{5} - \frac{1}{18} \right) = 1{,}1152.$$

Тогда $Q = \dfrac{1}{1{,}1152} \cdot \left(18 \cdot \ln 99{,}111 - \sum\limits_{j=1}^{4} f_j \cdot \ln s_j^* \right) = 0{,}692$.

Так как $Q = 0{,}692 < \chi^2_{0,95}(3) = 7{,}81$, гипотеза H_0' также не отклоняется.

4.1.3. Сравнение параметров биномиальных распределений

4.1.3.1. Сравнение двух параметров

Рассмотрим следующую задачу. В выборках объемов n_1 и n_2 из двух больших партий изделий зафиксировано соответственно m_1 и m_2 дефектных изделий. Требуется установить значимость различия долей дефектных приборов в партиях (p_1 и p_2), что тождественно проверке нулевой гипотезы $H_0: p_1 = p_2$ против возможных альтернатив H_1; $p_1 \neq p_2$, $H_1': p_1 > p_2$ и $H_1'': p_1 < p_2$.

При нормальной аппроксимации биномиального распределения (см. раздел 1.2.1)

статистикой для проверки H_0 является величина

$$z = \frac{\dfrac{m_1}{n_1} + \dfrac{1}{2n_1} - \dfrac{m_2}{n_2} - \dfrac{1}{2n_2}}{\sqrt{\dfrac{m_1+m_2}{n_1+n_2}\dfrac{n_1+n_2-m_1-m_2}{n_1+n_2}\left(\dfrac{1}{n}+\dfrac{1}{n_2}\right)}}.$$

При справедливости нулевой гипотезы z имеет стандартное нормальное распределение. Если $|z| > U_{\frac{1+\alpha}{2}}$, $z > U_\alpha$ или $z < U_{1-\alpha}$, нулевая гипотеза $H_0\colon p_1 = p_2$ отклоняется в пользу альтернатив

$$H_1\colon p_1 \neq p_2; \quad H_1'\colon p_1 > p_2 \quad \text{и} \quad H''\colon p_1 < p_2$$

соответственно (α — доверительная вероятность).

При использовании нормальной аппроксимации с помощью преобразования arcsin (см. раздел 1.2.1) статистика z имеет вид

$$z = \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\left(2\arcsin\sqrt{\frac{m_1}{n_1}} - 2\arcsin\sqrt{\frac{m_2}{n_2}}\right).$$

Задача 226. *В двух партиях приборов объемами $n_1 = 100$ шт. и $n_2 = 200$ шт. обнаружены соответственно $m_1 = 3$ и $m_2 = 5$ дефектных приборов. Необходимо проверить нулевую гипотезу о равенстве долей дефектных приборов в партиях при доверительной вероятности $\alpha = 0{,}95$.*

Вычисляем

$$z = \frac{\dfrac{3}{100} + \dfrac{1}{2\cdot 100} - \dfrac{5}{200} - \dfrac{1}{2\cdot 200}}{\sqrt{\dfrac{3+5}{100+200}\cdot\dfrac{100+200-3-5}{100+200}\cdot\left(\dfrac{1}{100}+\dfrac{1}{200}\right)}} = 0{,}376.$$

Из табл. 1 находим $u_{\frac{1+0{,}95}{2}} = u_{0{,}975} = 1{,}96$.

Так как $z = 0{,}376 < u_{0{,}975} = 1{,}96$, нулевая гипотеза не отклоняется.

Аналогично получаем

$$z = \sqrt{\frac{1}{100}+\frac{1}{200}}\cdot\left(2\arcsin\sqrt{\frac{3}{100}} - 2\arcsin\sqrt{\frac{5}{200}}\right) = 0{,}37,$$

что дает тот же результат.

4.1.3.2. Сравнение значения параметра с заданным

На практике доля дефектных изделий в партии является, как правило, критерием качества изделий на стадии производства. Основная задача выборочного контроля качества изделий заключается в том, чтобы по результатам проверки выборки изделий из партии установить, превосходит ли доля дефектных изделий в партии некоторую заранее нормированную контрольную величину. В терминах математической статистики такая задача формулируется как проверка нулевой гипотезы $H_0\colon p = p_0$ (здесь p — неизвестная истинная доля дефектных изделий в партии; p_0 — допустимая величина доли дефектных изделий в партии) против альтернатив $H_1\colon p > p_0$ или $H_1'\colon p < p_0$.

При контроле качества продукции представляет интерес, как правило, гипотеза $H_0\colon p \leqslant p_0$ (отклонение этой гипотезы альтернативой $H_1\colon p > p_0$ приводит к браковке партии изделий).

Статистикой для проверки гипотезы $H_0\colon p \leqslant p_0$ по выборке изделий объема n является величина

$$z = \frac{\frac{m}{n} - \frac{1}{2n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}},$$

имеющая при справедливости нулевой гипотезы стандартное нормальное распределение. Гипотеза $H_0\colon p \leqslant p_0$ отклоняется в пользу альтернативы $H_1\colon p > p_0$, если $z > u_\alpha$, где u_α — α-квантиль стандартного нормального распределения.

При использовании нормализующего преобразования arcsin значения z подсчитываются по формуле

$$z = \frac{2}{\sqrt{n}} \left(\arcsin \sqrt{\frac{m}{n}} - \arcsin \sqrt{p_0} \right),$$

где n — объем выборки, m — число дефектных изделий в выборке.

Задача 227. *Нормируемый уровень дефектных изделий в партии $p_0 = 0,05$. Предположим, что из партии изделий извлечена выборка $n = 20$ изделий, в которой обнаружены при проверке $m = 2$ дефектных изделия. Проверить с достоверностью $\alpha = 0{,}95$ гипотезу о том, что доля дефектных изделий в партии не превосходит нормируемого значения.*

Находим

$$z = \frac{\frac{2}{20} - \frac{1}{40} - 0{,}05}{\sqrt{\frac{0{,}05 \cdot (1 - 0{,}05)}{20}}} = 0{,}51.$$

Так как $z = 0{,}51 < u_{0,95} = 1{,}645$, проверяемая гипотеза не отклоняется.

4.1.3.3. Сравнение нескольких параметров ($k \geqslant 2$)

Предположим, имеются k выборок из биномиальных совокупностей, объемами $n_1, n_2, \ldots, n_k$. В каждой i-й выборке изучаемое событие наблюдалось m_i раз $(i = 1, \ldots, k)$. Необходимо проверить гипотезу $H_0\colon p_i = p_{i+1}$ $(i = 1, \ldots, k-1)$ против альтернативы $H_1\colon p_i \neq p_{i+1}$.

Статистикой критерия проверки H_0 является величина

$$\chi^2 = \frac{1}{\bar{p}\,(1-\bar{p})} \sum_{i=1}^{k} n_i \left(p_i - \bar{p}\right)^2, \quad \text{где} \quad \bar{p} = \frac{\sum\limits_{i=1}^{k} m_i}{\sum\limits_{i=1}^{k} n_i}; \quad p_i = \frac{m_i}{n_i}.$$

При справедливости нулевой гипотезы величина χ^2 имеет распределение хи-квадрат (см. раздел 1.1.8) с $f = k - 1$ степенями свободы. Поэтому если $\chi^2 > \chi^2_\alpha(k-1)$, где χ^2_α — α-квантиль распределения хи-квадрат, то с достоверностью α нулевая гипотеза отклоняется.

Если использовать нормализующее преобразование arcsin, то статистика критерия вычисляется по формуле

$$\chi^2 = \sum_{i=1}^{k} n_i \left(y_i - \bar{y}\right)^2, \quad \text{где} \quad y_i = 2 \arcsin \sqrt{\frac{m_i}{n_i}}, \quad \bar{y} = \frac{\sum\limits_{i=1}^{k} y_i}{\sum\limits_{i=1}^{k} n_i}.$$

Задача 228. *Из 8 различных партий изделий извлечены выборки объемами*

$$n_1 = 10, \quad n_2 = 20, \quad n_3 = 10, \quad n_4 = 50, \quad n_5 = 30, \quad n_6 = 25, \quad n_7 = 10, \quad n_8 = 40.$$

При проверке в них были обнаружены следующие количества дефектных изделий:

$$m_1 = 2, \quad m_2 = 5, \quad m_3 = 3, \quad m_4 = 8, \quad m_5 = 10, \quad m_6 = 6, \quad m_7 = 3, \quad m_8 = 7.$$

Необходимо проверить гипотезу о равенстве долей дефектных изделий в партиях изделий с достоверностью $\alpha = 0{,}99$.

Имеем

$$p_1 = \frac{2}{20} = 0{,}2; \quad p_2 = \frac{5}{20} = 0{,}4; \quad p_3 = \frac{3}{10} = 0{,}3; \quad p_4 = \frac{8}{50} = 0{,}16;$$

$$p_5 = \frac{10}{30} = 0{,}33; \quad p_6 = \frac{6}{25} = 0{,}24; \quad p_7 = \frac{3}{10} = 0{,}3; \quad p_8 = \frac{7}{40} = 0{,}175.$$

Вычисляем

$$\bar{p} = \frac{\sum\limits_{i=1}^{8} m_i}{\sum\limits_{i=1}^{8} n_i} = \frac{2+5+3+8+10+6+3+7}{10+20+10+50+30+25+10+40} = 0{,}2256.$$

Далее

$$\chi^2 = \frac{1}{0{,}2256 \cdot (1 - 0{,}2256)} \cdot \sum_{i=1}^{8} n_i \cdot (p_i - \bar{p})^2 = 5{,}7239 \cdot \Big[10 \cdot (0{,}2 - 0{,}2256)^2 +$$
$$+ 20 \cdot (0{,}4 - 0{,}2256)^2 + 10 \cdot (0{,}3 - 0{,}2256)^2 + 50 \cdot (0{,}16 - 0{,}2256)^2 + 30 \cdot (0{,}33 - 0{,}2256)^2 +$$
$$+ 25 \cdot (0{,}24 - 0{,}2256)^2 + 10 \cdot (0.33 - 0{,}2256)^2 + 40 \cdot (0{,}175 - 0{,}2256)^2\Big] = 8{,}325.$$

При $\alpha = 0{,}99$ имеем $\chi^2_{0,99}(f = k - 1 = 7) = 18{,}48$.

Так, как $\chi^2 = 8{,}325 < \chi^2_{0,99}(7) = 18{,}48$, нулевая гипотеза не отклоняется.

4.1.4. Последовательные методы проверки гипотез о значениях параметров распределений (последовательный анализ Вальда)

Рассмотренные ранее методы проверки статистических гипотез предполагали фиксированный объем выборки. Вальд [433] предложил теорию последовательной проверки гипотез (последовательный анализ), существенным отличием которой является то, что число наблюдений, необходимое для принятия решения по гипотезе, зависит от исходов испытаний и является случайной, не фиксированной заранее величиной.

Метод последовательной проверки гипотезы предполагает на каждой стадии наблюдений (эксперимента) принятие одного из возможных решений: принять гипотезу, отклонить ее или продолжить наблюдения. Обычно при последовательном анализе нулевая гипотеза относительно значения параметра ε формулируется в форме предположения об одном из двух его возможных значений ε_0 или ε_1:

$$H_0\colon \varepsilon = \varepsilon_0, \quad H_1\colon \varepsilon = \varepsilon_1.$$

Задача последовательного анализа в ходе эксперимента — выбрать одну из гипотез. Вальд [433] показал, что для проверки гипотез методами последовательного анализа требуется в среднем в два раза меньше наблюдений, чем при проверке классическими методами, основанными на заранее фиксированном числе наблюдений. Впоследствии было показано [434], что при определенных условиях выигрыш от применения последовательной процедуры по сравнению с классической теоретически неограничен.

Перед планированием процедуры последовательного анализа назначают приемлемые величины вероятностей допустимых ошибок: α — вероятность принятия гипотезы H_1, когда верна гипотеза H_0 (ошибка первого рода) и β — вероятность принятия гипотезы H_0, когда верна гипотеза H_1 (ошибка второго рода).

Наибольший выигрыш последовательный анализ дает при $\alpha \gg \beta$ или $\alpha \ll \beta$, т. е. когда α и β являются величинами разного порядка малости.

Так как число наблюдений n в последовательном анализе является величиной случайной, то необходимо знать либо его функцию распределения вероятностей, либо параметры этого распределения (например, среднее количество необходимых наблюдений). Среднее значение зависит только от истинного значения параметра, относительно которого проверяется гипотеза.

Функция $\bar{n}(\varepsilon)$, определяющая зависимость $\bar{n}$ от ε, называется функцией среднего числа наблюдений.

На практике обычно находят средние значения числа наблюдений $\bar{n}(\varepsilon_0)$ и $\bar{n}(\varepsilon_1)$ соответствующие гипотетическим значениям параметра ε_0 и ε_1, между которыми осуществляется выбор, и максимальное среднее значение числа наблюдений $\bar{n}_{\max}$, необходимое для окончания последовательной процедуры проверки гипотезы.

Если знания только среднего числа наблюдений недостаточно и требуется определить либо вероятность того, что для завершения последовательной процедуры понадобится не более некоторого, наперед заданного, числа наблюдений, либо число наблюдений, соответствующее заданной вероятности завершения последовательной процедуры, используются таблицы распределения Вальда [435] с функцией

$$p\,(x < a) = W_c(a) = \sqrt{\frac{c}{2\pi}} \int\limits_0^a x^{-\frac{3}{2}} \exp\left\{-\frac{c}{2}\left(x + \frac{1}{x} - 2\right)\right\} dx,$$

где $x = \dfrac{n}{\bar{n}}$ — отношение числа наблюдений к его среднему значению; c — параметр распределения, определяемый видом распределения исследуемой случайной величины и гипотетическим значением параметра (ε_0 или ε_1).

Вероятность γ завершения процедуры последовательного анализа и число испытаний $n(\varepsilon)$ при некотором значении ε (ε_0 или ε_1) связаны соотношением $\gamma = W_{c(\varepsilon)} \dfrac{n\,(\varepsilon)}{\bar{n}\,(\varepsilon)}$, из которого можно определить либо γ, соответствующее заданному $n(\varepsilon)$, либо $n(\varepsilon)$, соответствующее заданному значению γ.

Некоторые, наиболее употребляемые значения функции распределения Вальда $W_c\left(x = \dfrac{n}{\bar{n}}\right)$, приведены в табл. 142.

4.1.4.1. Проверка гипотез о параметрах нормального распределения

4.1.4.1.1. Проверка гипотезы о значении среднего

Проверяется нулевая гипотеза $H_0\colon \mu = \mu_0$ против альтернативы $H_1\colon \mu = \mu_1$ ($\mu_0 < \mu_1$). Полагается, что дисперсии σ_1^2 и σ_2^2 известны заранее, причем $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Введем обозначения

$$A = \frac{\sigma^2}{\mu_1 - \mu_0} \ln \frac{\beta}{1-\alpha} + n\frac{\mu_0 + \mu_1}{2};$$

$$B = \frac{\sigma^2}{\mu_1 - \mu_0} \ln \frac{1-\beta}{\alpha} + n\frac{\mu_0 + \mu_1}{2}.$$

Таблица 142

Функция распределения $W_c(x)$ Вальда [435]

x	c							
	0,01	0,02	0,03	0,04	0,05	0,07	0,10	0,20
0,006	0,1986	0,0692	0,0261	0,0102	0,0041			
0,007	0,2343	0,0928	0,0396	0,0175	0,0079	0,0017		
0,008	0,2662	0,1161	0,0544	0,0264	0,0130	0,0033		
0,009	0,2948	0,1388	0,0699	0,0364	0,0194	0,0057		
0,01	0,3205	0,1605	0,0858	0,0473	0,0266	0,0087	0,0017	
0,02	0,4843	0,3237	0,2227	0,1637	0,1196	0,0658	0,0280	0,0019
0,03	0,5693	0,4225	0,3269	0,2582	0,2067	0,1357	0,0749	0,0120
0,04	0,6232	0,4891	0,3981	0,3301	0,2769	0,1992	0,1256	0,0309
0,05	0,6612	0,5376	0,4518	0,3860	0,3334	0,2536	0,1736	0,0554
0,06	0,6899	0,5750	0,4939	0,4309	0,3795	0,3000	0,2170	0,0826
0,07	0,7125	0,6048	0,5281	0,4677	0,4282	0,3399	0,2559	0,1106
0,08	0,7309	0,6294	0,5565	0,4987	0,4508	0,3744	0,2906	0,1383
0,09	0,7465	0,6500	0,5806	0,5252	0,4790	0,4046	0,3217	0,1652
0,10	0,7593	0,6677	0,6013	0,5482	0,5036	0,4313	0,3498	0,1909
0,11	0,7706	0,6831	0,6195	0,5683	0,5252	0,4555	0,3751	0,2153
0,12	0,7805	0,6967	0,6355	0,5682	0,5445	0,4764	0,3981	0,2385
0,13	0,7893	0,7087	0,6498	0,6022	0,5619	0,4957	0,4191	0,2604
0,14	0,7971	0,7194	0,6626	0,6166	0,5775	0,5132	0,4384	0,2810
0,15	0,8041	0,7292	0,6742	0,6297	0,5918	0,5292	0,4592	0,3006
0,20	0,8312	0,7667	0,7192	0,6806	0,6476	0,5928	0,5277	0,3835
0,25	0,8498	0,7926	0,7506	0,7163	0,6870	0,6381	0,5797	0,4478
0,30	0,8635	0,8119	0,7740	0,7431	0,7166	0,6725	0,6196	0,4990
0,35	0,8743	0,8270	0,7923	0,7641	0,7400	0,6997	0,6514	0,5409
0,40	0,8829	0,8391	0,8071	0,7811	0,7589	0,7219	0,6776	0,5760
0,45	0,8901	0,8493	0,8194	0,7953	0,7748	0,7405	0,6996	0,5958
0,50	0,8962	0,8578	0,8300	0,8074	0,7882	0,7563	0,7183	0,6316
0,55	0,9015	0,8562	0,8390	0,8178	0,7998	0,7700	0,7345	0,6541
0,60	0,9060	0,8717	0,8469	0,8269	0,8100	0,7819	0,7488	0,6739
0,65	0,9101	0,8774	0,8590	0,8350	0,8190	0,7926	0,7614	0,6916
0,70	0,9137	0,8824	0,8601	0,8421	0,8270	0,8020	0,7727	0,7074
0,75	0,9169	0,8871	0,8657	0,8486	0,8342	0,8106	0,7829	0,7218
0,80	0,9198	0,8912	0,8707	0,8544	0,8407	0,8183	0,7921	0,7347
0,85	0,9225	0,8949	0,8754	0,8598	0,8467	0,8253	0,8005	0,7466
0,90	0,9249	0,8984	0,8796	0,8646	0,8521	0,8318	0,8082	0,7575
0,95	0,9272	0,9016	0,8835	0,8691	0,8571	0,8377	0,8153	0,7675
1,0	0,9292	0,9045	0,8871	0,8771	0,8618	0,8419	0,8219	0,7768
1,1	0,9329	0,9098	0,8835	0,8807	0,8701	0,8530	0,8336	0,7934
1,2	0,9362	0,9143	0,8991	0,8871	0,8773	0,8616	0,8439	0,8097
1,3	0,9390	0,9183	0,9040	0,8929	0,8873	0,8691	0,8529	0,8206
1,4	0,9416	0,9220	0,9074	0,8980	0,8894	0,8758	0,8608	0,8319
1,5	0,9349	0,9252	0,9124	0,9025	0,8945	0,8818	0,8681	0,8419
1,6	0,9459	0,9281	0,9160	0,9066	0,8991	0,8873	0,8745	0,8510
1,7	0,9478	0,9308	0,9192	0,9104	0,9033	0,8922	0,8804	0,8592
1,8	0,9495	0,9332	0,9222	0,9138	0,9071	0,8967	0,8858	0,8667
1,9	0,9511	0,9354	0,9249	0,9170	0,9106	0,9009	0,8907	0,8735
2,0	0,9526	0,9375	0,9274	0,9199	0,9138	0,9047	0,8952	0,8798
2,5	0,9585	0,9458	0,9376	0,9316	0,9269	0,9201	0,9134	0,9046
3,0	0,9629	0,9520	0,9451	0,9402	0,9365	0,9313	0,9266	0,9222
3,5	0,9663	0,9568	0,9509	0,9468	0,9438	0,9398	0,9366	0,9353

Продолжение таблицы 142

x	c							
	0,01	0,02	0,03	0,04	0,05	0,07	0,10	0,20
4,0	0,9690	0,9606	0,9555	0,9521	0,9497	0,9466	0,9445	0,9454
4,5	0,9713	0,9638	0,9594	0,9565	0,9545	0,9522	0,9508	0,9538
5,0	0,9732	0,9663	0,9626	0,9601	0,9585	0,9568	0,9561	0,9598
6,0	0,9762	0,9706	0,9677	0,9659	0,9649	0,9640	0,9643	0,9694
7,0	0,9786	0,9739	0,9716	0,9703	0,9697	0,9694	0,9704	0,9762
8,0	0,9805	0,9765	0,9747	0,9738	0,9734	0,9737	0,9750	0,9812
9,0	0,9820	0,9786	0,9772	0,9766	0,9765	0,9771	0,9787	0,9849
10,0	0,9833	0,9804	0,9793	0,9790	0,9790	0,9798	0,9816	0,9878
15,0	0,9877	0,9863	0,9862	0,9865	0,9870	0,9884	0,9904	0,9952
20,0	0,9903	0,9897	0,9900	0,9906	0,9912	0,9926	0,9944	0,9979
25,0	0,9920	0,9918	0,9924	0,9931	0,9938	0,9951	0,9966	0,9990
30,0	0,9932	0,9932	0,9940	0,9947	0,9954	0,9966	0,9980	0,9996
40,0	0,9949	0,9954	0,9961	0,9968	0,9974	0,9982	0,9991	0,9999
50,0	0,9960	0,9966	0,9973	0,9979	0,9984	0,9990	0,9995	0,9999
100,0	0,9983	0,9990	0,9994	0,9996	0,9998	0,9999	0,9999	0,9999
x	0,3	0,5	0,7	0,9	1,0	1,2	1,4	1,6
0,03	0,0021							
0,04	0,0083							
0,05	0,0192	0,0025						
0,06	0,0340	0,0063	0,0012					
0,07	0,0515	0,0122	0,0031					
0,08	0,0706	0,0201	0,0069					
0,09	0,0907	0,0298	0,0104					
0,10	0,1112	0,0410	0,0159	0,0064	0,0041	0,0017		
0,11	0,1316	0,0533	0,0227	0,0090	0,0066	0,0030		
0,12	0,1518	0,0664	0,0306	0,0145	0,0100	0,0049	0,0024	0,0012
0,13	0,1715	0,0802	0,0395	0,0199	0,0143	0,0074	0,0039	0,0020
0,14	0,1907	0,0945	0,0491	0,0262	0,0193	0,0106	0,0058	0,0032
0,15	0,2093	0,1089	0,0595	0,0333	0,0251	0,0144	0,0083	0,0049
0,20	0,2926	0,1814	0,1175	0,0778	0,0637	0,0431	0,0295	0,0202
0,25	0,3614	0,2492	0,1787	0,1309	0,1127	0,0842	0,0634	0,0481
0,30	0,4183	0,3039	0,2379	0,1863	0,1657	0,1322	0,1062	0,0859
0,35	0,4662	0,3635	0,2930	0,2406	0,2190	0,1829	0,1538	0,1301
0,40	0,5070	0,4109	0,3435	0,2921	0,2706	0,2337	0,2033	0,1777
0,45	0,5421	0,4528	0,3894	0,3402	0,3194	0,2832	0,2527	0,2266
0,50	0,5727	0,4901	0,4310	0,3878	0,3650	0,3303	0,3007	0,2750
0,55	0,5997	0,5234	0,4688	0,4258	0,4072	0,3746	0,3466	0,3220
0,60	0,6236	0,5534	0,5031	0,4635	0,4464	0,4162	0,3900	0,3670
0,65	0,6450	0,5804	0,5343	0,4981	0,4825	0,4548	0,4308	0,4096
0,70	0,6643	0,6048	0,5628	0,5299	0,5158	0,4907	0,4690	0,4498
0,75	0,6817	0,6271	0,5889	0,5592	0,5464	0,5239	0,5045	0,4874
0,80	0,6976	0,6474	0,6128	0,5861	0,5747	0,5547	0,5376	0,5225
0,85	0,7121	0,6661	0,6348	0,6109	0,6008	0,5833	0,5683	0,5552
0,90	0,7254	0,6833	0,6551	0,6339	0,6250	0,6097	0,5967	0,5855
1,0	0,7490	0,7138	0,6911	0,6748	0,6681	0,6568	0,6477	0,6400
1,1	0,7693	0,7400	0,7221	0,7099	0,7025	0,6975	0,6915	0,6868
1,2	0,7870	0,7628	0,7491	0,7404	0,7372	0,7325	0,7293	0,7272
1,3	0,8025	0,7828	0,7726	0,7669	0,7651	0,7629	0,7619	0,7619
1,4	0,8163	0,8003	0,7932	0,7901	0,7894	0,7893	0,7902	0,7928

Продолжение таблицы 142

x	c							
	0,3	0,5	0,7	0,9	1,0	1,2	1,4	1,6
1,5	0,8286	0,8160	0,8114	0,8105	0,8108	0,8123	0,8148	0,8177
1,6	0,8396	0,8299	0,8276	0,8284	0,8295	0,8325	0,8361	0,8401
1,7	0,8495	0,8424	0,8420	0,8487	0,8461	0,8502	0,8547	0,8595
1,8	0,8585	0,8536	0,8549	0,8585	0,8608	0,8658	0,8701	0,8764
1,9	0,8667	0,8638	0,8664	0,8711	0,8738	0,8795	0,8853	0,8911
2,0	0,8742	0,8731	0,8769	0,8824	0,8855	0,8917	0,8979	0,9039
2,5	0,9036	0,9086	0,9161	0,9240	0,9278	0,9350	0,9416	0,9476
3,0	0,9240	0,9321	0,9411	0,9434	0,9532	0,9599	0,9657	0,9707
3,5	0,9389	0,9486	0,9578	0,9656	0,9690	0,9748	0,9794	0,9833
4,0	0,9500	0,9604	0,9692	0,9762	0,9791	0,9838	0,9875	0,9903
4,5	0,9587	0,9690	0,9772	0,9833	0,9857	0,9895	0,9922	0,9943
5,0	0,9654	0,9755	0,9830	0,9881	0,9901	0,9931	0,9952	0,9966
6,0	0,9753	0,9844	0,9902	0,9939	0,9951	0,9969	0,9981	0,9988
7,0	0,9819	0,9898	0,9942	0,9967	0,9975	0,9986	0,9992	0,9995
8,0	0,9865	0,9932	0,9965	0,9982	0,9987	0,9993	0,9996	0,9997
9,0	0,9898	0,9954	0,9979	0,9990	0,9993	0,9997	0,9998	0,9998
10,0	0,9922	0,9968	0,9987	0,9994	0,9996	0,9998	0,9999	0,9999
15,0	0,9977	0,9994	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999
20,0	0,9993	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
50,0	0,9996	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
x	1,8	2,0	2,2	2,4	2,6	2,8	3,0	3,2
0,15	0,0028	0,0017						
0,20	0,0140	0,0097	0,0068	0,0047	0,0033	0,0023	0,0016	0,0011
0,25	0,0366	0,0280	0,0215	0,0166	0,0128	0,0099	0,0770	0,0060
0,30	0,0697	0,0569	0,0465	0,0382	0,0314	0,0259	0,0214	0,0177
0,35	0,1105	0,0942	0,0806	0,0691	0,0594	0,0511	0,0441	0,0381
0,40	0,1561	0,1375	0,1215	0,1076	0,0955	0,0849	0,0756	0,0674
0,45	0,2039	0,1842	0,1667	0,1513	0,0138	0,1253	0,1143	0,1044
0,50	0,2524	0,2323	0,2144	0,1983	0,1837	0,1705	0,1584	0,1474
0,55	0,3000	0,2806	0,2629	0,2468	0,2320	0,2185	0,2060	0,1945
0,60	0,3464	0,3278	0,3109	0,2953	0,2810	0,2678	0,2554	0,2440
0,65	0,3906	0,3734	0,2576	0,3431	0,3296	0,3170	0,3053	0,2943
0,70	0,4325	0,4168	0,4025	0,3892	0,3769	0,3654	0,3545	0,3443
0,75	0,4720	0,4580	0,4452	0,4334	0,4223	0,4121	0,4024	0,3932
0,80	0,5100	0,4967	0,4856	0,4752	0,4657	0,4567	0,4483	0,4403
0,85	0,5435	0,5330	0,5235	0,5147	0,5065	0,4990	0,4919	0,4852
0,90	0,5757	0,5669	0,5589	0,5516	0,5450	0,5387	0,5330	0,5376
0,95	0,6056	0,5984	0,.5920	0,5861	0,5808	0,5759	0,5714	0,5673
1,0	0,6334	0,6277	0,6227	0,6182	0,6142	0,6106	0,6073	0,6043
1,1	0,6831	0,6800	0,6776	0,6755	0,6738	0,6725	0,6713	0,6704
1,2	0,7258	0,7250	0,7246	0,7245	0,7247	0,7252	0,6958	0,7265
1,3	0,7624	0,7634	0,7647	0,7662	0,7679	0,7698	0,7717	0,7737
1,4	0,7939	0,7936	0,7989	0,8016	0,8044	0,8073	0,8102	0,8131
1,5	0,8201	0,8244	0,8280	0,8316	0,8352	0,8388	0,8423	0,8458
1,6	0,8442	0,8484	0,8527	0,8569	0,8611	0,8651	0,8691	0,8729
1,7	0,8643	0,8691	0,8738	0,8784	0,8828	0,8872	0,8913	0,8953
1,8	0,8816	0,8818	0,8918	0,8966	0,9011	0,9056	0,9098	0,9138
1,9	0,9066	0,9020	0,9071	0,9120	0,9166	0,9209	0,9251	0,9290
2,0	0,9096	0,9150	0,9218	0,9250	0,9295	0,9338	0,9378	0,9415

Продолжение таблицы 142

x	c							
	1,8	2,0	2,2	2,4	2,6	2,8	3,0	3,2
2,5	0,9530	0,9578	0,9621	0,9659	0,9694	0,9725	0,9725	0,9777
3,0	0,9749	0,9785	0,9816	0,9842	0,9865	0,9884	0,9900	0,9914
3,5	0,9863	0,9889	0,9909	0,9926	0,9939	0,9950	0,9959	0,9966
4,0	0,9925	0,9942	0,9954	0,9964	0,9972	0,9978	0,9983	0,9987
4,5	0,9958	0,9969	0,9977	0,9983	0,9987	0,9990	0,9992	0,9994
5,0	0,9976	0,9983	0,9988	0,9992	0,9994	0,9996	0,9997	0,9998
6,0	0,9992	0,9995	0,9997	0,9998	0,9999	0,9999	0,9999	0,9999
7,0	0,9997	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
8,0	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
x	3,4	3,6	3,8	4,0	4,2	4,4	4,6	4,8
0,25	0,0046	0,0036	0,0028	0,0022	0,0017	0,0013	0,0010	
0,30	0,0146	0,0121	0,0100	0,0084	0,0070	0,0058	0,0048	
0,35	0,0330	0,0285	0,0247	0,0214	0,0186	0,0162	0,0141	
0,40	0,0602	0,0538	0,0481	0,0431	0,0386	0,0347	0,0311	
0,45	0,0955	0,0874	0,0801	0,0735	0,0674	0,0619	0,0569	
0,50	0,1373	0,1280	0,1195	0,1116	0,1043	0,0975	0,0913	
0,55	0,1838	0,1738	0,1645	0,1559	0,1478	0,1402	0,1330	
0,60	0,2332	0,2231	0,2136	0,2046	0,1962	0,1882	0,1806	
0,65	0,2839	0,2741	0,2648	0,2560	0,2477	0,2397	0,2321	
0,70	0,3347	0,3256	0,3169	0,3086	0,3007	0,2931	0,2859	
0,75	0,3846	0,2763	0,3685	0,3610	0,3539	0,3470	0,3404	
0,80	0,4328	0,4256	0,4188	0,4123	0,4061	0,4000	0,3943	
0,85	0,4789	0,4729	0,4671	0,4617	0,4565	0,4515	0,4467	
0,90	0,5225	0,5176	0,5131	0,5087	0,5046	0,5006	0,4968	
0,95	0,5634	0,5597	0,5563	0,5530	0,5499	0,5470	0,5442	
1,0	0,6015	0,6000	0,5966	0,5944	0,5923	0,5904	0,5886	
1,1	0,6700	0,6691	0,6686	0,6682	0,6680	0,6678	0,6677	0,6677
1,2	0,7274	0,7283	0,7296	0,7305	0,7316	0,7328	0,7341	0,7353
1,3	0,7758	0,7779	0,7800	0,7822	0,7844	0,7865	0,7887	0,7909
1,4	0,8160	0,8189	0,8218	0,8247	0,8275	0,8303	0,8330	0,8357
1,5	0,8493	0,8527	0,8560	0,8593	0,8624	0,8656	0,8686	0,8716
1,6	0,8767	0,8804	0,8839	0,8873	0,8906	0,8939	0,9870	0,9000
1,7	0,8992	0,9029	0,9065	0,9099	0,9132	0,9164	0,9194	0,9224
1,8	0,9176	0,9213	0,9248	0,9281	0,9312	0,9343	0,9371	0,9399
1,9	0,9327	0,9362	0,9395	0,9426	0,9455	0,9484	0,9510	0,9535
2,0	0,9450	0,9483	0,9514	0,9542	0,9570	0,9595	0,9619	0,9641
2,5	0,9799	0,9819	0,9837	0,9853	0,9868	0,9881	0,9892	0,9903
3,0	0,9926	0,9936	0,9945	0,9953	0,9959	0,9965	0,9970	0,9974
4,0	0,9969	0,9992	0,9994	0,9995	0,9996	0,9997	0,9998	0,9998
5,0	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
x	5,0	5,5	6,0	6,5	7,0	7,5	8,0	8,5
0,30	0,0033	0,0021	0,0014					
0,35	0,0107	0,0076	0,0054	0,0039	0,0028	0,0020	0,0014	0,0010
0,40	0,0251	0,0192	0,0148	0,0115	0,0089	0,0069	0,0053	0,0041
0,45	0,0482	0,0392	0,0320	0,0261	0,0215	0,0176	0,0145	0,0120
0,50	0, 0801	0,0682	0,0582	0,0498	0,0427	0,0366	0,0315	0,0271
0,55	0,1200	0,1058	0,0935	0,0828	0,0735	0,0653	0,0581	0,0518
0,60	0,1666	0,1510	0,1371	0,1248	0,1137	0,1038	0,0949	0,0869

Окончание таблицы 142

x	c							
	5,0	5,5	6,0	6,5	7,0	7,5	8,0	8,5
0,65	0,2180	0,2019	0,1874	0,1742	0,1623	0,1513	0,1412	0,1320
0,70	0,2723	0,2567	0,2424	0,2293	0,2172	0,2060	0,1956	0,1859
0,75	0,3279	0,3135	0,3002	0,2879	0,2765	0,2658	0,2557	0,2462
0,80	0,3834	0,3707	0,3590	0,3481	0,3379	0,3283	0,3193	0,3107
0,85	0,4376	0,4270	0,4173	0,4082	0,3997	0,3917	0,3841	0,3769
0,90	0,4897	0,4814	0,4738	0,4668	0,4602	0,4541	0,4482	0,4429
0,95	0,5390	0,5331	0,5277	0,5228	0,5283	0,5140	0,5101	0,5053
1,1	0,6677	0,6680	0,6685	0,6693	0,6702	0,6712	0,6723	0,6734
1,2	0,7366	0,7399	0,7433	0,7467	0,7501	0,7535	0,7569	0,7603
1,3	0,7930	0,7983	0,8036	0,8086	0,8136	0,8184	0,8231	0,8277
1,4	0,8384	0,8449	0,8511	0,8570	0,8627	0,8681	0,8733	0,8782
1,5	0,8745	0,8815	0,8880	0,8941	0,8999	0,9053	0,9104	0,9152
1,6	0,9030	0,9099	0,9163	0,9222	0,9276	0,9326	0,9473	0,9416
1,7	0,9252	0,9318	0,9377	0,9431	0,9480	0,9524	0,9565	0,9601
1,8	0,9425	0,9485	0,9539	0,9586	0,9629	0,9666	0,9700	0,9730
1,9	0,9559	0,9613	0,9659	0,9700	0,9736	0,9767	0,9795	0,9819
2,0	0,9662	0,9709	0,9749	0,9784	0,9813	0,9838	0,9860	0,9879
2,5	0,9912	0,9932	0,9947	0,9959	0,9968	0,9975	0,9980	0,9985
3,0	0,9977	0,9984	0,9989	0,9992	0,9995	0,9996	0,9997	0,9997
4,0	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
5,0	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
x	9,0	10	15	20	25	30	35	40
0,40	0,0032	0,0020						
0,50	0,0234	0,0174	0,0042	0,0010				
0,55	0,0462	0,0369	0,0124	0,0045				
0,60	0,0796	0,0670	0,0294	0,0134	0,0063	0,0030	0,0014	
0,65	0,1235	0,1084	0,0584	0,0320	0,0131	0,0064	0,0064	0,0032
0,70	0,1768	0,1603	0,1015	0,0864	0,0442	0,0131	0,0204	0,0140
0,75	0,2373	0,2209	0,1586	0,1100	0,0840	0,0620	0,0512	0,0330
0,80	0,3026	0,2874	0,2279	0,1852	0,1528	0,1273	0,1068	0,0901
0,85	0,3701	0,3572	0,3057	0,2310	0,2311	0,2111	0,1873	0,1640
0,90	0,4374	0,4276	0,3881	0,3580	0,3334	0,3124	0,2940	0,2775
0,95	0,5028	0,4963	0,4709	0,4512	0,4284	0,4250	0,4120	0,3940
1,0	0,5648	0,5616	0,5507	0,5441	0,5395	0,5361	0,5335	0,5313
1,1	0,6747	0,6772	0,6913	0,7052	0,7184	0,7307	0,7421	0,7528
1,2	0,7636	0,7701	0,7996	0,8243	0,8452	0,8631	0,8795	0,8919
1,3	0,8321	0,8405	0,8758	0,9022	0,9223	0,9379	0,9501	0,9597
1,4	0,8830	0,8918	0,9260	0,9486	0,9638	0,9743	0,9817	0,9869
1,5	0,9197	0,9279	0,9573	0,9742	0,9842	0,9902	0,9939	0,9961
1,6	0,9455	0,9457	0,9760	0,9876	0,9934	0,9965	0,9981	0,9990
1,7	0,9635	0,9693	0,9868	0,9942	0,9974	0,9988	0,9995	0,9997
1,8	0,9757	0,9802	0,9929	0,9974	0,9990	0,9996	0,9998	0,9999
1,9	0,9840	0,9874	0,9962	0,9988	0,9996	0,9999	0,9999	0,9999
2,0	0,9895	0,9921	0,9980	0,9995	0,9999	0,9999	0,9999	0,9999
2,5	0,9988	0,9993	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,0	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

Если:

$\sum\limits_{i=1}^{n} x_i \leqslant A$, то принимается гипотеза H_0;

$\sum\limits_{i=1}^{n} x_i \geqslant B$, то принимается гипотеза H_1;

$A < \sum\limits_{i=1}^{n} x_i < B$, то наблюдения продолжаются.

Средние объемы выборок, необходимые для завершения процедуры последовательного анализа, равны

$$\bar{n}(\mu_0) = 2\sigma^2 \frac{(1-\alpha)\ln\dfrac{1-\alpha}{\beta} + \alpha\ln\dfrac{\alpha}{1-\beta}}{(\mu_1-\mu_0)^2};$$

$$\bar{n}(\mu_1) = 2\sigma^2 \frac{\beta\ln\dfrac{\beta}{1-\alpha} + (1-\beta)\ln\dfrac{1-\beta}{\alpha}}{(\mu_1-\mu_0)^2};$$

$$\bar{n}_{\max} = -\sigma^2 \frac{\ln\dfrac{\beta}{1-\alpha}\ln\dfrac{1-\beta}{\alpha}}{(\mu_1-\mu_0)^2}.$$

Параметр c распределения Вальда находится по формуле

$$c = K\frac{\left|\mu - \dfrac{\mu_0+\mu_1}{2}\right|}{\mu_1-\mu_0}, \quad \text{где} \quad K = \begin{cases} \ln\dfrac{1-\alpha}{\beta}, & \text{если } \mu=\mu_0, \quad \alpha \ll \beta; \\ \ln\dfrac{1-\beta}{\alpha}, & \text{если } \mu=\mu_1, \quad \beta \ll \alpha. \end{cases}$$

Задача 229. *Предположим, что параметр прибора распределен нормально со стандартным отклонением $\sigma = 200$. Необходимо проверить при $\alpha = 0{,}1$ и $\beta = 0{,}01$ гипотезу о том, что параметр прибора равен $\mu = \mu_0 = 1800$, против альтернативы $\mu = \mu_1 = 2000$.*

Найти контрольные границы и средние объемы выборок для последовательной проверки гипотезы. Определить объем выборки $n_{0,95}$, для которого с вероятностью не менее 0,95 процедура последовательного анализа закончится принятием решения по гипотезе.

Находим

$$A = \frac{200^2}{200-1800}\cdot\ln\frac{0{,}01}{0{,}9} + n\cdot\frac{2000+1800}{2} = -899{,}96 + 1900\cdot n;$$

$$B = \frac{200^2}{2000-1800}\ln\frac{0{,}99}{0{,}1} + n\cdot\frac{2000+1800}{2} = 458{,}5 + 1900\cdot n.$$

Далее вычисляем средние объемы выборок (округляем до ближайшего большего целого):

$$\bar{n}\,(1800) = 2\cdot 200^2\cdot\frac{0{,}9\cdot\ln\dfrac{0{,}9}{0{,}01} + 0{,}1\cdot\ln\dfrac{0{,}1}{0{,}99}}{(2000-1800)^2} = 8;$$

$$\bar{n}\,(2000) = 2\cdot 200^2\cdot\frac{0{,}01\cdot\ln\dfrac{0{,}01}{0{,}9} + 0{,}99\cdot\ln\dfrac{0{,}99}{0{,}1}}{(2000-1800)^2} = 5;$$

$$\bar{n}_{\max} = -200\cdot\frac{\ln\dfrac{0{,}01}{0{,}9}\cdot\ln\dfrac{0{,}99}{0{,}1}}{(2000-1800)^2} = 10.$$

В нашем случае $\beta \ll \alpha$, тогда для $\mu_1 = 2000$ находим

$$K = \ln\frac{0{,}99}{0{,}1} = 2{,}2925; \quad c = 2{,}2925 \cdot \frac{\left|2000 - \dfrac{2000+1800}{2}\right|}{2000-1800} = 1{,}14625.$$

При $\mu = \mu_1$ мы ранее получили $\bar{n}(\mu_1) = 5$. Из табл. 142 для $c = 1{,}1462$ находим $x = \dfrac{n_{0,95}}{\bar{n}}$ соответствующее условию $W_c(x) = 0{,}95$.

Для $c = 1{,}1462$ и $\gamma = 0{,}95$ имеем (интерполируя) $x = \dfrac{n_{0,95}}{\bar{n}} = 2{,}3$.

Тогда $n_{0,95} = 2{,}3 \cdot \bar{n}(2000) = 2{,}3 \cdot 5 = 14$, т. е. с вероятностью 0,95 для принятия решения по гипотезе потребуется не более 14 испытаний.

Итак, принимаем гипотезу H_0, если

$$\sum_{i=1}^{m} x_i \leqslant -899{,}96 + 1900 \cdot n,$$

и принимаем гипотезу H_1, если

$$\sum_{i=1}^{n} x_i \geqslant 458{,}5 + 1900 \cdot n.$$

В любом ином случае испытания необходимо продолжить. При H_0 в среднем понадобятся 8 испытаний, а при H_1 — 5 испытаний. Максимальное среднее число испытаний не превысит 10.

4.1.4.1.2. Проверка гипотезы о значении дисперсии

Проверяется гипотеза $H_0\colon \sigma^2 = \sigma_0^2$ против альтернативы $H_1\colon \sigma^2 = \sigma_1^2$ $(\sigma_1 > \sigma_0)$ при известном среднем μ.

Гипотеза H_0 принимается, если $\sum\limits_{i=1}^{n} x_i^2 \leqslant A$; если $\sum\limits_{i=1}^{n} x_i^2 \geqslant B$, принимается гипотеза H_1; если $A < \sum\limits_{i=1}^{n} x_i^2 < B$, то испытания продолжаются. Здесь

$$A = \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\ln\left[\frac{\beta}{1-\alpha}\left(\frac{\sigma_1}{\sigma_0}\right)^n\right]; \quad B = \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\ln\left[\frac{1-\beta}{\alpha}\left(\frac{\sigma_1}{\sigma_0}\right)^n\right].$$

Средние объемы выборок, необходимые для завершения последовательной процедуры, равны

$$\bar{n}(\sigma_0^2) = 2\frac{(1-\alpha)\ln\dfrac{1-\alpha}{\beta} + \alpha\ln\dfrac{\alpha}{1-\beta}}{2\ln\dfrac{\sigma_1}{\sigma_0} - 1 + \left(\dfrac{\sigma_0}{\sigma_1}\right)^2}; \quad \bar{n}(\sigma_1^2) = 2\frac{\beta\ln\dfrac{\beta}{1-\alpha} + (1-\beta)\ln\dfrac{1-\beta}{\alpha}}{\left(\dfrac{\sigma_1}{\sigma_0}\right)^2 - 1 - 2\ln\dfrac{\sigma_1}{\sigma_0}};$$

$$\bar{n}_{\max} = -\frac{2\ln\dfrac{\beta}{1-\alpha}\ln\dfrac{1-\beta}{\alpha}}{\ln^2\left(\dfrac{\sigma_1}{\sigma_0}\right)^2}.$$

Вероятностные оценки необходимого числа испытаний могут быть найдены, по аналогии с критерием для проверки гипотезы о среднем значении, с помощью табл. 142 распределения Вальда при

$$c = K\frac{\left|\sigma^2 - \dfrac{\sigma_0^2 + \sigma_1^2}{2}\right|}{\sigma_1^2 - \sigma_0^2}, \quad \text{где} \quad K = \begin{cases} \ln\dfrac{1-\alpha}{\beta}, & \text{если } \sigma^2 = \sigma_0^2, \quad \alpha \ll \beta; \\ \ln\dfrac{1-\beta}{\alpha}, & \text{если } \sigma^2 = \sigma_1^2, \quad \beta \ll \alpha. \end{cases}$$

Если значение среднего μ неизвестно, то все приведенные соотношения сохраняются при замене n на $(n-1)$.

Задача 230. *Предположим, что параметр прибора распределен нормально с известным средним μ. Необходимо проверить при $\alpha = 0{,}01$ и $\beta = 0{,}1$ гипотезу о том, что дисперсия значений параметра прибора равна $\sigma^2 = \sigma_0^2 = 25$ против альтернативы, утверждающей, что $\sigma^2 = \sigma_1^2 = 49$.*

Найти контрольные границы и средние объемы выборок для последовательной процедуры проверки гипотез. Определить объем выборки $n_{0,9}$, для которого с вероятностью не менее $\gamma = 0{,}90$ процедура последовательного анализа закончится принятием решения по гипотезе. Определить вероятность того, что последовательная процедура потребует не более 20 испытаний.

Находим

$$A = \frac{2\cdot 25\cdot 49}{49-25}\cdot \ln\left[\frac{0{,}1}{0{,}99}\cdot\left(\frac{7}{5}\right)^n\right] = -234{,}029 + 34{,}348\cdot n;$$

$$B = \frac{2\cdot 25\cdot 49}{49-25}\cdot \ln\left[\frac{0{,}9}{0{,}01}\cdot\left(\frac{7}{3}\right)^n\right] = 459{,}354 + 34{,}348\cdot n.$$

Если $\sum\limits_{i=1}^{n} x_i^2 \leqslant -234,029 + 34,348\cdot n$, то принимается гипотеза о том, что $\sigma^2 = 25$; если $\sum\limits_{i=1}^{n} x_i^2 \geqslant 459{,}354 + 34{,}348\cdot n$, то принимается гипотеза о том, что $\sigma^2 = 49$. В случае $-234{,}029 + 34{,}348\cdot n < \sum\limits_{i=1}^{n} x_i^2 < 459{,}354 + 34{,}348\cdot n$ испытания необходимо продолжить.

Далее вычисляем средние объемы выборок (округляя до ближайшего целого числа)

$$\bar{n}(25) = 2\cdot\frac{0{,}99\cdot\ln\dfrac{0{,}1}{0{,}99} + 0{,}01\cdot\ln\dfrac{0{,}01}{0{,}9}}{2\cdot\ln\dfrac{7}{5} - 1 + \left(\dfrac{5}{7}\right)^2} = 25;\quad \bar{n}(49) = 2\cdot\frac{0{,}1\cdot\ln\dfrac{0{,}1}{0{,}99} + 0{,}9\cdot\ln\dfrac{0{,}9}{0{,}01}}{\left(\dfrac{7}{5}\right)^2 - 1 - 2\cdot\ln\dfrac{7}{5}} = 27;$$

$$\bar{n}_{\max} = -\frac{2\cdot\ln\dfrac{0{,}1}{0{,}99}\cdot\ln\dfrac{0{,}9}{0{,}01}}{\ln^2\left(\dfrac{7}{5}\right)^2} = 46.$$

В нашем случае $\alpha \ll \beta$, тогда для $\sigma^2 = \sigma_0^2 = 25$ находим

$$K = \left|\ln\frac{0{,}1}{0{,}99}\right| = 2{,}292 \quad \text{и} \quad c = 2{,}292\cdot\frac{\left|25 - \dfrac{25+49}{2}\right|}{49-25} = 1{,}146.$$

Из табл. 142 имеем $W_{1,146}(x) = 0{,}90$ при $x = \dfrac{n_{0,9}}{\bar{n}} \approx 2{,}8$, тогда $n_{0,9} = \bar{n}\left(\sigma_0^2\right)\cdot 2{,}8 = 70$. Теперь определим вероятность окончания последовательной процедуры при $n \leqslant 20$. Имеем $x = \dfrac{n}{\bar{n}} = \dfrac{20}{25} = 0{,}8$.

Из табл. 142 находим для $c = 1{,}146$: $W_{1,146}(x) = 0{,}56$.

Следовательно, с вероятностью 0,56 для окончания последовательной процедуры потребуется не более 20 испытаний, при условии, что $\sigma^2 = \sigma_0^2 = 25$.

4.1.4.2. Проверка гипотезы о параметре экспоненциального распределения

Проверяется гипотеза $H_0\colon \lambda = \lambda_0$ (напомним, что $\lambda = 1/\nu$ — средняя наработка до отказа, ν — интенсивность отказов), против альтернативы $H_1\colon \lambda = \lambda_1 > \lambda_0$.

Введем обозначения

$$A = \frac{n}{\lambda_1 - \lambda_0} \ln\left[\frac{\lambda_1}{\lambda_0}\left(\frac{\alpha}{1-\beta}\right)^{\frac{1}{n}}\right]; \quad B = \frac{n}{\lambda_1 - \lambda_0} \ln\left[\frac{\lambda_1}{\lambda_0}\left(\frac{1-\alpha}{\beta}\right)^{\frac{1}{n}}\right].$$

Если $\sum_{i=1}^{n} x_i \leqslant A$, принимается гипотеза H_0; если $\sum_{i=1}^{n} x_i \geqslant B$, принимается гипотеза H_1; если $A < \sum_{i=1}^{n} x_i < B$, испытания продолжаются.

Средние объемы выборок для последовательной процедуры определяются по формулам

$$\bar{n}(\lambda_0) = \frac{(1-\alpha)\ln\dfrac{1-\alpha}{\beta} + \alpha\ln\dfrac{\alpha}{1-\beta}}{\dfrac{\lambda_1}{\lambda_0} - 1 - \ln\dfrac{\lambda_1}{\lambda_0}}; \quad \bar{n}(\lambda_1) = \frac{\beta\ln\dfrac{\beta}{1-\alpha} + (1-\beta)\ln\dfrac{1-\beta}{\alpha}}{\dfrac{\lambda_0}{\lambda_1} - 1 + \ln\dfrac{\lambda_1}{\lambda_0}};$$

$$\bar{n}_{\max} = -\frac{\ln\left(\dfrac{1-\alpha}{\beta}\right)\ln\left(\dfrac{\alpha}{1-\beta}\right)}{\ln^2\left(\dfrac{\lambda_1}{\lambda_0}\right)}.$$

Параметр распределения Вальда находится по формуле

$$c = K\frac{\left|\ln\dfrac{\lambda_1}{\lambda_0} - \dfrac{\lambda_1 - \lambda_0}{2}\right|}{\left(\dfrac{\lambda_1 - \lambda_0}{2}\right)^2}, \quad \text{где} \quad K = \begin{cases} \ln\dfrac{1-\alpha}{\beta}, & \text{если } \lambda = \lambda_0, \quad \alpha \ll \beta; \\ \ln\dfrac{1-\beta}{\alpha}, & \text{если } \lambda = \lambda_1, \quad \beta \ll \alpha. \end{cases}$$

Задача 231. *Найти параметры плана последовательных испытаний при $\alpha = 0{,}2$ и $\beta = 0{,}05$ для проверки гипотезы H_0 о том, что средняя наработка на отказ электронного прибора равна $\lambda = \lambda_0 = 100$ ч против альтернативы $H_1: \lambda = \lambda_1 = 150$ ч. Вычислить средние объемы выборок, необходимых для окончания последовательного анализа. Найти вероятность того, что последовательная процедура позволит принять решение уже после $n = 20$ испытаний.*

Находим

$$A = \frac{n}{150 - 100} \cdot \ln\left[\frac{150}{100} \cdot \left(\frac{0{,}2}{0{,}95}\right)^{\frac{1}{n}}\right] = 0{,}00811 \cdot n - 0{,}0312;$$

$$B = \frac{n}{150 - 100} \cdot \ln\left[\frac{150}{100} \cdot \left(\frac{0{,}8}{0{,}05}\right)^{\frac{1}{n}}\right] = 0{,}00811 \cdot n + 0{,}0554.$$

Далее вычисляем:

$$\bar{n}(100) = \frac{0{,}8 \cdot \ln\dfrac{0{,}8}{0{,}05} + 0{,}2 \cdot \ln\dfrac{0{,}2}{0{,}95}}{\dfrac{150}{100} - 1 + \ln\dfrac{150}{100}} = 21;$$

$$\bar{n}(150) = \frac{0{,}05 \cdot \ln\dfrac{0{,}05}{0{,}8} + 0{,}95 \cdot \ln\dfrac{0{,}95}{0{,}2}}{\dfrac{100}{150} - 1 + \ln\dfrac{50}{100}} = 19; \quad \bar{n}_{\max} = -\frac{\ln\dfrac{0{,}8}{0{,}05} \cdot \ln\dfrac{0{,}2}{0{,}95}}{\ln^2\dfrac{150}{100}} = 27.$$

Теперь найдем, учитывая, что $\beta \ll \alpha$ и $\lambda = \lambda_1 = 150$,

$$K = \ln\frac{0{,}95}{0{,}2} = 1{,}558 \quad \text{и} \quad c = 1{,}558 \cdot \frac{\left|\ln\frac{150}{100} - \frac{150-100}{2}\right|}{\left(\frac{150-100}{150}\right)^2} = 344{,}864.$$

Имеем $x = \dfrac{n}{\bar{n}\,(\lambda_1)} = \dfrac{20}{19} \approx 1$ и из табл. 142 для $x = 1$ и $c = 344$ имеем $W_c(x) = 0{,}53$, т. е. с вероятностью 0,53 процедура последовательного анализа закончится при $n = 20$, если $\lambda = \lambda_1 = 150$.

4.1.4.3. Проверка гипотезы о параметре биномиального распределения

Проверяется гипотеза $H_0\colon p = p_0$ против альтернативы $H_1\colon p = p_1 > p_0$.

Введем обозначения

$$A = \frac{\ln\dfrac{\beta}{1-\alpha}}{\ln\dfrac{p_1}{p_0} + \ln\dfrac{1-p_0}{1-p_1}} + n\frac{\ln\dfrac{1-p_0}{1-p_1}}{\ln\dfrac{p_1}{p_0} + \ln\dfrac{1-p_0}{1-p_1}};$$

$$B = \frac{\ln\dfrac{1-\beta}{\alpha}}{\ln\dfrac{p_1}{p_0} + \ln\dfrac{1-p_0}{1-p_1}} + n\frac{\ln\dfrac{1-p_0}{1-p_1}}{\ln\dfrac{p_1}{p_0} + \ln\dfrac{1-p_0}{1-p_1}}.$$

Пусть x — число наступлений наблюдаемого события (например, количество дефектных приборов в партии). Если $x \leqslant A$, то принимается гипотеза H_0; если $x \geqslant B$, принимается гипотеза H_1; в случае $A < x < B$ испытания продолжаются.

Средние объемы выборок равны

$$\bar{n}(p_0) = \frac{(1-\alpha)\ln\dfrac{\beta}{1-\alpha} + \alpha\ln\dfrac{1-\beta}{\alpha}}{p_0\ln\dfrac{p_1}{p_0} - (1-p_0)\ln\dfrac{1-p_0}{1-p_1}};$$

$$\bar{n}(p_1) = \frac{\beta\ln\dfrac{\beta}{1-\alpha} + (1-\beta)\ln\dfrac{1-\beta}{\alpha}}{p_1\ln\dfrac{p_1}{p_0} - (1-p_1)\ln\dfrac{1-p_0}{1-p_1}};$$

$$\bar{n}_{\max} = \frac{\ln\left(\dfrac{1-\beta}{\alpha}\right)\ln\left(\dfrac{\beta}{1-\alpha}\right)}{\ln\dfrac{p_1}{p_0}\ln\dfrac{1-p_0}{1-p_1}}.$$

Параметр распределения Вальда находится по формуле

$$c = K\frac{\left|p\ln\dfrac{p_1}{p_0} - (1-p)\ln\dfrac{1-p_0}{1-p_1}\right|}{p\,(1-p)\left(\ln\dfrac{p_1}{p_0} + \ln\dfrac{1-p_0}{1-p_1}\right)},$$

где

$$K = \begin{cases} \ln\dfrac{1-\alpha}{\beta}, & \text{если } p = p_0, \quad \alpha \ll \beta; \\ \ln\dfrac{1-\beta}{\alpha}, & \text{если } p = p_1, \quad \beta \ll \alpha. \end{cases}$$

Задача 232. *Необходимо найти параметры плана последовательных испытаний для проверки гипотезы H_0 о том, что доля дефектных изделий в партии $p = p_0 = 0{,}01$ против альтернативы $H_1: p = p_1 = 0{,}02$. Определить количество испытаний, для которого с вероятностью $\gamma = 0{,}95$ последовательная процедура закончится принятием решения (заданы $\alpha = 0{,}1$ и $\beta = 0{,}01$).*

Находим

$$A = \frac{\ln\dfrac{0{,}01}{0{,}9}}{\ln\dfrac{0{,}02}{0{,}01} + \ln\dfrac{0{,}99}{0{,}98}} + n\cdot\frac{\ln\dfrac{0{,}99}{0{,}98}}{\ln\dfrac{0{,}02}{0{,}01} + \ln\dfrac{0{,}99}{0{,}98}} = -6{,}399 + 0{,}0144\cdot n;$$

$$B = \frac{\ln\dfrac{0{,}99}{0{,}1}}{\ln\dfrac{0{,}02}{0{,}01} + \ln\dfrac{0{,}99}{0{,}98}} + n\cdot\frac{\ln\dfrac{0{,}99}{0{,}98}}{\ln\dfrac{0{,}02}{0{,}01} + \ln\{0{,}99\}\,0{,}98} = 3{,}26 + 0{,}0144\cdot n.$$

Далее

$$\bar{n}(0{,}01) = \frac{0{,}9\cdot\ln\dfrac{0{,}01}{0{,}9} + 0{,}1\cdot\ln\dfrac{0{,}99}{0{,}1}}{0{,}01\cdot\ln\dfrac{0{,}01}{0{,}9} - 0{,}99\cdot\ln\dfrac{0{,}99}{0{,}98}} = 1225; \quad \bar{n}(0{,}02) = \frac{0{,}01\cdot\ln\dfrac{0{,}01}{0{,}9} + 0{,}99\cdot\ln\dfrac{0{,}99}{0{,}1}}{0{,}02\cdot\ln\dfrac{0{,}02}{0{,}01} - 0{,}98\cdot\ln\dfrac{0{,}99}{0{,}98}} = 569;$$

$$\bar{n}_{\max} = -\frac{\ln\dfrac{0{,}99}{0{,}1}\cdot\ln\dfrac{0{,}01}{0{,}9}}{\ln\dfrac{0{,}02}{0{,}01}\cdot\ln\dfrac{0{,}99}{0{,}98}} = 1466.$$

Вычисляем параметр c (при $\beta \ll \alpha$ и $p = p_1$):

$$K = \ln\frac{1-\beta}{\alpha} = \ln\frac{0{,}99}{0{,}1} = 2{,}2925 \quad \text{и} \quad c = 2{,}2925\cdot\frac{\left|0{,}02\cdot\ln\dfrac{0{,}02}{0{,}01} - 0{,}98\cdot\ln\dfrac{0{,}99}{0{,}98}\right|}{0{,}02\cdot 0{,}98\cdot\left(\ln\dfrac{0{,}02}{0{,}01} + \ln\dfrac{0{,}99}{0{,}98}\right)} = 0{,}65.$$

Теперь из табл. 142 находим $x = \dfrac{n_{0,95}}{\bar{n}}$, соответствующее условию $W_c(x) = 0{,}95$. Имеем $W_{0,65}(x) = 0{,}95$ при $x \approx 3{,}5$, или $n_{0,95} = 569\cdot 3{,}5 = 1992$.

Следовательно, с вероятностью 0,95 при $p = p_1 = 0{,}02$ процедура последовательного анализа потребует не более 1992 испытаний.

4.2. Непараметрические (свободные от распределения) критерии однородности статистических данных

Рассмотренные ранее методы сравнения параметров распределений предполагали, что мы заранее обладаем фундаментальной информацией — нам известен вид закона распределения вероятностей.

Обычно это требование редко является препятствием для желающих применять статистику для решения своих проблем. Однако отклонение закона распределения вероятностей изучаемой случайной величины от нам понравившегося приводит к искажению вероятностных характеристик наших выводов (вплоть до принятия решения, противоположного правдоподобному).

В тех случаях, когда наши предположения о гипотетическом законе распределения вероятностей не кажутся убедительными, следует применять иные методы для сравнения случайных величин и проверки гипотез об их значениях.

Методы статистической обработки и анализа результатов наблюдений, закон распределения вероятностей появления которых неизвестен, объединены единым направлением математической статистики, получившим название *непараметрическая статистика*. Ее приемы и методы, известные еще как *методы, свободные от распределения*, интенсивно развиваются в последние годы.

Любое распределение можно описать *параметром положения*, характеризующим центр группирования случайных величин, и *параметром масштаба*, характеризующим степень рассеяния случайных величин относительно центра группирования (например, в случае нормального распределения ими являются соответственно среднее μ и стандартное отклонение σ).

Когда закон распределения неизвестен, гипотезы о параметрах положения и масштаба проверяются с помощью *специальных критериев сдвига и масштаба*.

Если $f_1(x)$ и $f_2(x)$ — неизвестные плотности вероятностей, то гипотеза сдвига записывается как $H_0\colon f_1(x) = f_2(x)$ против альтернативы $H_1\colon f_1(x) = f_2(x-\Delta)$, или $H_0\colon \Delta = 0$ против альтернативы $H_1\colon \Delta \neq 0$, где Δ — сдвиг, определяемый разностью параметров положения распределений.

Гипотезы о разнице в дисперсиях (при неизвестных распределениях) формулируются как гипотезы о параметрах масштаба.

Например, если

$$f_1(x) = \frac{1}{\tau} f\left(\frac{x-\mu}{\tau}\right) \quad \text{и} \quad f_2(x) = f(x-\mu),$$

то гипотеза о параметре масштаба записывается как $H_0\colon \tau = 1$ против альтернативы $H_1\colon \tau \neq 1$.

Достоинством рассматриваемых в настоящем разделе непараметрических (свободных от распределения) методов проверки статистических гипотез является их расчетная простота. Однако мощность статистических критериев, построенных на их основе, уступает аналогичным параметрическим критериям (например, критериям Стьюдента, Фишера и т. п.). Легко догадаться, что это плата за незнание вида распределения случайных величин.

Рекомендуется следующий порядок использования непараметрических критериев. Если распределение случайной величины неизвестно, то непараметрические критерии являются единственно возможными критериями для проверки различных статистических гипотез. Если распределение известно, то рекомендуется сначала применить простые в вычислительном отношении непараметрические критерии. При отклонении ими проверяемой гипотезы дальнейшее уточнение не требуется. Если непараметрический критерий не отклоняет гипотезу, необходимо осуществить ее дальнейшую проверку одним из более точных параметрических критериев, изложенных в разделе 4.1.

4.2.1. Непараметрические критерии сдвига

4.2.1.1. Сравнение параметров сдвига двух совокупностей

4.2.1.1.1. Быстрый (грубый) критерий Кенуя

Критерий изложен в [121]. Алгоритм его построения состоит в следующем.

Пусть $0 < \varkappa < 1$. Через $x_\varkappa$ обозначим такое значение случайной величины, которое превышается $[\varkappa n]$ значениями из выборки.

Вычисляется среднее значение

$$m = 0{,}2x_{\frac{1}{16}} + 0{,}6x_{\frac{1}{2}} + 0{,}2x_{\frac{15}{16}}$$

со стандартным отклонением

$$\frac{1{,}1s}{\sqrt{n}}, \quad \text{где} \quad s = \frac{1}{3}\left(x_{\frac{1}{16}} - x_{\frac{15}{16}}\right).$$

Для очень несимметричных распределений используются оценки

$$m = \frac{1}{6}\left(x_{\frac{1}{16}} + x_{\frac{1}{4}} + 2x_{\frac{1}{2}} + x_{\frac{3}{4}} + x_{\frac{15}{16}}\right); \quad s = \frac{1}{4}\left(x_{\frac{1}{16}} + \frac{3}{4}x_{\frac{1}{4}} - \frac{3}{4}x_{\frac{3}{4}} - x_{\frac{15}{16}}\right).$$

Проверка разности в параметрах положения проверяется критерием

$$M = \frac{m_1 - m_2}{\sqrt{s_1^2 + s_2^2}},$$

(индексы относятся к номерам проверяемых выборок).

При объемах проверяемых выборок свыше 20 статистика критерия распределена нормально. Поэтому нулевая гипотеза отсутствия сдвига не отклоняется при доверительной вероятности α, если $|M| < u_{\frac{1+\alpha}{2}}$.

Критерий устойчив к отклонениям от нормальности, имеет эффективность по сравнению с параметрическим критерием Стьюдента не хуже $\approx 93\%$.

Задача 233. *Имеются две группы наблюдений объема $n = 32$:*

(гр. 1):	1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11,	12,	13,	14,	15,	16,
	17,	18,	19,	20,	21,	22,	23,	24,	25,	26,	27,	28,	29,	30,	31,	32;
(гр. 2):	10,	11,	12,	13,	14,	15,	16,	17,	18,	19,	20,	21,	22,	23,	24,	25,
	26,	27,	28,	29,	30,	31,	32,	33,	34,	35,	36,	37,	38,	39,	40,	41.

Необходимо проверить отсутствие сдвига между средними критерием Кенуя при доверительной вероятности $\alpha = 0{,}95$.

Находим для первой совокупности

$$x_{\frac{1}{16}} = 30; \quad x_{\frac{1}{2}} = 16; \quad x_{\frac{15}{16}} = 2; \quad m_1 = 0{,}2 \cdot 30 + 0{,}6 \cdot 16 + 0{,}2 \cdot 2 = 16;$$

$$s_1 = \frac{30 - 2}{3} = 9{,}33.$$

Для второй совокупности имеем

$$x_{\frac{1}{16}} = 39; \quad x_{\frac{1}{2}} = 25; \quad x_{\frac{15}{16}} = 11; \quad m_2 = 0{,}2 \cdot 39 + 0{,}6 \cdot 25 + 0{,}2 \cdot 11 = 25;$$

$$s_2 = \frac{39 - 11}{3} = 9{,}33.$$

Далее

$$M = \frac{|m_1 - m_2|}{\sqrt{s_1^2 + s_2^2}} = \frac{9}{13{,}199} = 0{,}682.$$

Для $\alpha = 0{,}95$ имеем $u_{\frac{1+0{,}95}{2}} = u_{0{,}975} = 1{,}96$.

Так как $|M| = 0{,}682 < u_{0{,}975} = 1{,}96$, гипотеза сдвига отклоняется.

Рассмотрим второй вариант критерия. Для первой выборки имеем

$$x_{\frac{1}{4}} = 28; \quad x_{\frac{3}{4}} = 4; \quad m_1 = \frac{x_{\frac{1}{16}} + x_{\frac{1}{4}} + 2 \cdot x_{\frac{1}{2}} + x_{\frac{3}{4}} + x_{\frac{15}{16}}}{6} = \frac{30 + 28 + 2 \cdot 16 + 4 + 2}{6} = 16;$$

$$s_1 = \frac{1}{4} \cdot \left(30 + \frac{3}{4} \cdot 28 - \frac{3}{4} \cdot 4 - 2\right) = 11{,}5.$$

Аналогично для второй выборки имеем

$$x_{\frac{1}{4}} = 37; \quad x_{\frac{3}{4}} = 13; \quad m_2 = 25; \quad s_2 = 11{,}5.$$

Получаем $M = 0{,}555$; легко видеть, что в силу симметричности наших выборок оба метода дают одинаковые результаты.

4.2.1.1.2. Ранговые критерии сдвига

Ранговые критерии основываются на последовательности рангов выборочных значений случайных величин. При этом рассматриваются не сами выборочные значения, а их ранги, определяемые порядковым номером элемента выборки в общем ряду, упорядоченном по возрастанию. Например, в упорядоченной выборке $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ выборочное значение x_i заменяется рангом $R = i$.

4.2.1.1.2.1. Быстрый (грубый) ранговый критерий

Изложен в [121]. Рассматриваются две выборки объемов n и m при $n + m = 20$ ($n, m \geqslant 4$). Их элементы ранжируются по возрастанию совместно. Одинаковым наблюдениям присваивается одинаковый усредненный ранг. Для каждой группы находятся суммы рангов $\sum R_1$ и $\sum R_2$ и средние ранги $\bar{R}_1 = \frac{1}{n}\sum R_i$ и $\bar{R}_2 = \frac{1}{m}\sum R_j$.

Вычисляем $d = R_1 - R_2$. Статистика d-критерия может быть аппроксимирована нормальным распределением со средним $\mathbf{M}(d) = 0$ и дисперсией

$$s_d = \sqrt{\frac{\sum R_1 + \sum R_2}{6}\left(\frac{1}{m} + \frac{1}{n}\right)} = \sqrt{\frac{(m+n)(m+n+1)}{12}\left(\frac{1}{m} + \frac{1}{n}\right)}.$$

Поэтому при $|d^*| = \left|\frac{d}{s_d}\right| < u_{\frac{1+\alpha}{2}}$ гипотеза сдвига отклоняется с доверительной вероятностью α.

Эффективность критерия для нормально распределенных выборок 0,95 (для любого другого исходного распределения — не хуже 0,86).

Задача 234. *Имеются две выборки случайных величин* ($n = 12$, $m = 8$):

x_{1i}: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12;
x_{2i}: 10, 11, 12, 13, 14, 15, 16, 17.

Необходимо проверить гипотезу сдвига быстрым ранговым критерием при доверительной вероятности $\alpha = 0{,}95$.

Совместный ранжированный ряд имеет вид (вверху проставлены значения случайной величины, внизу — принадлежность к выборке):

1	2	3	4	5	6	7	8	9	10	10	11	11	12	12	13	14	15	16	17
x_1	x_1	x_1	x_1	x_1	x_1	x_1	x_1	x_1	x_1	x_2	x_1	x_2	x_1	x_2	x_2	x_2	x_2	x_2	x_2

Ранги элементов первой выборки (x_{1i}) равны (одинаковые ранги усредняются) 1; 2; 3; 4; 5; 6; 7; 8; 9; 10,5; 11,5; 12,5 и $\sum R_1 = 79{,}5$; $\bar{R}_1 = 6{,}625$.

Для второй выборки имеем ряд

$$R_2\colon 10{,}5;\ 11{,}5;\ 12{,}5;\ 13;\ 14;\ 15;\ 16;\ 17 \quad \text{и} \quad \sum R_2 = 109{,}5; \quad \bar{R}_2 = 13{,}687.$$

Далее $d = \left|\bar{R}_1 - \bar{R}_2\right| = |6{,}625 - 13{,}687| = 7{,}062$; $|d^*| = \dfrac{7{,}062}{2{,}70} = 2{,}61$. Так как $|d^*| = 2{,}61 > u_{\frac{1+\alpha}{2}} = 1{,}96$, гипотеза сдвига не отклоняется.

4.2.1.1.2.2. Критерий Манна–Уитни–Вилкоксона

Пусть $x_1, \ldots, x_n$ и $y_1, \ldots, y_m$ — упорядоченные по возрастанию выборки. Для проверки гипотезы сдвига Манн и Уитни [436] предложили ранговый критерий, основанный на статистике

$$U = \sum_{i=1}^{n}\sum_{j=1}^{m} h_{ij}, \quad \text{где} \quad h_{ij} = \begin{cases} 1, & x_i < y_j; \\ 0, & x_i > y_j. \end{cases}$$

Здесь U — точное число пар значений x_i и y_j, для которых $x_i < y_j$.

Если $U_1(\alpha) \leqslant U \leqslant U_2(\alpha)$, гипотеза сдвига отклоняется ($U_1(\alpha)$ и $U_2(\alpha)$ — критические значения, приведенные в табл. 143).

С U-статистикой Манна–Уитни связана статистика Вилкоксона [437], определяемая суммой рангов элементов одной выборки (предположим, x_i объема n) в общей упорядоченной последовательности элементов совместной выборки объема $(m+n)$:

$$R = mn + \frac{n\,(n+1)}{2} - U.$$

При $n, m > 20$ применима аппроксимация

$$W = \frac{R - \dfrac{n\,(n+m+1)}{2}}{\sqrt{\dfrac{nm\,(n+m+1)}{12}}}.$$

Статистика W аппроксимируется нормальным распределением, и гипотеза сдвига отклоняется с достоверностью α, если $|W| > u_{\frac{1+\alpha}{2}}$.

Если в двух сравниваемых выборках есть совпадающие значения, то им рекомендуется приписывать средние ранги (среднеарифметическое для каждой серии последовательных рангов). При этом в знаменателе статистики следует использовать величину

$$\left\{\frac{nm(n+m+1)}{12}\left[1 - \frac{\sum\limits_{i=1}^{k} t_i(t_i^2-1)}{(m+n)(m+n-1)(m+n+1)}\right]\right\}^{\frac{1}{2}},$$

где k — общее число групп совпадающих величин; t_i — число совпавших величин в i-й группе (следует помнить, что совпадения учитываются только тогда, когда совпавшие величины принадлежат различным выборкам, т. е. совпадения, целиком состоящие из элементов одной и той же выборки, на величину W не влияют).

Таблица 143

Критические значения $U_1(\alpha)$ и $U_2(\alpha)$ критерия Манна–Уитни

(α — доверительная вероятность) [57]

n	m	α				n	m	α			
		0,90		0,95				0,90		0,95	
		U_1	U_2	U_1	U_2			U_1	U_2	U_1	U_2
4	4	1	15	0	16	10	13	37	93	33	97
	5	2	18	1	19		14	41	99	36	104
	6	3	21	2	22		15	44	106	39	111
	7	4	24	3	25		16	48	112	42	118
	8	5	27	4	28		17	51	119	45	125
	9	6	30	4	32		18	55	125	48	132
	10	7	33	5	35		19	58	132	52	138
5	5	4	21	2	23		20	62	138	55	145
	6	5	25	3	27	12	12	42	102	37	107
	7	6	29	5	30		13	47	109	41	117
	8	8	32	6	34		14	51	117	45	123
	9	9	36	7	38		15	55	125	49	131
	10	11	39	8	42		16	60	132	53	139
6	6	7	29	5	31		17	64	140	57	147
	7	8	34	6	36		18	68	148	61	155
	8	10	38	8	40		19	72	156	65	163
	9	12	42	10	44		20	77	163	69	171
	10	14	46	11	49	14	14	61	135	55	141
	11	16	50	13	53		15	66	144	59	151
	12	17	55	14	58		16	71	153	64	160
7	7	11	38	8	41		17	77	161	69	169
	8	13	43	10	46		18	82	170	74	178
	9	15	48	12	51		19	87	179	78	188
	10	17	53	14	56		20	92	188	83	197
	11	19	58	16	61		21	97	197	88	206
	12	21	63	18	66		22	102	206	93	215
	13	24	67	20	71	16	16	83	173	75	181
	14	26	72	22	76		17	89	183	81	191
8	8	15	49	13	51		18	95	193	86	202
	9	18	57	15	54		19	101	203	92	212
	10	20	60	17	63		20	107	213	98	222
	11	23	65	19	69		21	113	223	103	233
	12	26	70	22	74		22	119	233	109	243
	13	28	76	24	80		23	125	243	115	253
	14	31	81	26	86		24	131	253	120	264
	15	33	87	29	91	18	18	109	215	99	225
	16	36	92	31	97		19	116	226	106	236
9	9	21	60	17	64		20	123	237	112	248
	10	24	66	20	70		21	130	248	119	259
	11	27	72	23	76		22	136	260	125	271
	12	30	78	26	82		23	143	271	132	282
	13	33	84	28	89		24	150	282	138	294
	14	36	90	31	95		25	157	293	145	305
	15	39	96	34	101		26	164	304	151	317
	16	42	102	37	107	20	20	138	262	127	273
10	10	27	73	23	77		21	146	274	134	286
	11	31	79	26	84		22	154	286	141	299
	12	34	86	29	91		23	161	299	149	311

Окончание таблицы 143

n	m	α				n	m	α			
		0,90		0,95				0,90		0,95	
		U_1	U_2	U_1	U_2			U_1	U_2	U_1	U_2
20	24	169	311	156	324	28	32	347	549	315	571
	25	177	323	163	337		33	359	565	326	598
	26	185	335	171	349		34	370	582	337	615
	27	192	348	178	362	30	30	338	562	317	583
	28	200	360	186	374		31	350	580	328	602
22	22	171	313	158	326		32	362	598	340	620
	23	179	327	166	340		33	374	616	352	638
	24	188	340	174	354		34	387	633	364	656
	25	197	353	182	368		35	399	651	375	675
	26	205	367	191	381		36	411	669	387	693
	27	214	380	199	395	32	32	388	636	365	659
	28	223	393	207	409		33	402	654	378	678
	29	231	407	215	423		34	415	673	391	697
	30	240	420	223	437		35	428	692	403	717
24	24	207	369	192	384		36	441	711	416	736
	25	217	383	201	399		37	454	730	428	756
	26	226	398	210	414		38	467	749	441	775
	27	236	412	219	429	34	34	443	713	418	738
	28	245	427	228	444		35	457	733	431	759
	29	255	441	238	458		36	471	753	445	779
	30	264	456	247	473		37	485	773	458	800
	31	274	470	256	488		38	499	793	472	820
	32	284	484	265	503		39	513	813	485	841
26	26	247	429	230	446		40	527	833	499	861
	27	257	445	240	462	36	36	471	753	445	779
	28	268	460	250	478		37	486	774	459	801
	28	278	476	260	494		38	500	795	473	822
	30	289	491	270	510		39	515	815	487	843
	31	299	507	280	526		40	529	836	501	864
	32	310	522	290	542	38	38	563	881	533	911
28	28	291	493	272	512		39	578	904	548	934
	29	302	510	282	530		40	594	926	563	957
	30	313	527	293	547	40	40	628	972	596	1004
	31	325	543	304	564						

Более точная аппроксимация предложена Иманом [438]. В соответствии с ней гипотеза сдвига отклоняется с достоверностью α', если

$$|J| =< J(\alpha'),$$

где

$$J = \frac{W}{2}\left\{1 + \left(\frac{n+m-2}{n+m-1-W^2}\right)^{\frac{1}{2}}\right\}; \quad J(\alpha') = \frac{1}{2z_{\alpha'}} + \frac{1}{2t_{\alpha'}(f)};$$

$z_{\alpha'}$ — α'-квантиль нормального распределения; $t_{\alpha'}(f)$ — α'-квантиль распределения Стьюдента с $f = n+m-2$ степенями свободы; $\alpha' = \dfrac{1+\alpha}{2}$.

Асимптотическая эффективность критерия Манна–Уитни равна $3/\pi \approx 0{,}95$.

Одним из вариантов применения рассмотренного критерия является так называемый ранговый критерий Вилкоксона. Его статистика строится следующим образом. Для двух выборок x и y одинакового объема n строится ряд разностей $|x_i - y_i|$, который затем ранжируется по возрастанию.

В упорядоченном ряду значений $|x_i - y_i|$ находится сумма рангов T величин $z_i = x_i - y_i > 0$. Гипотеза сдвига отклоняется, если $T_1(\alpha) \leqslant T \leqslant T_2(\alpha)$, где $T_1(\alpha)$ и $T_2(\alpha)$ — критические значения, приведенные в табл. 144.

Таблица 144

Критические значения статистики T знакового рангового критерия Вилкоксона (α — доверительная вероятность) [57]

n	α				n	α				n	α			
	0,90		0,95			0,90		0,95			0,90		0,95	
	T_1	T_2	T_1	T_2		T_1	T_2	T_1	T_2		T_1	T_2	T_1	T_2
6	2	19	0	21	13	22	69	18	73	20	61	149	53	157
7	4	24	3	25	14	26	79	22	83	22	76	177	66	187
8	6	30	4	32	15	31	89	26	94	24	92	208	82	218
9	9	36	6	39	16	36	100	30	106	26	111	240	99	252
10	11	44	9	46	17	42	111	35	118	28	131	275	117	289
11	14	51	11	55	18	48	123	41	130	30	152	313	138	327
12	18	60	14	64	19	54	136	47	143	32	176	353	160	368

При $n \geqslant 20$ применимо приближение

$$T^* = \frac{T - \dfrac{n\,(n+1)}{4}}{\sqrt{n\,(n+1)\,(2n+1)}}.$$

При $|T^*| < u_{\frac{1+\alpha}{2}}$ гипотеза сдвига отклоняется (здесь u_γ — γ-квантиль стандартного нормального распределения).

Более точно приближение содержится в [439], в соответствии с ним гипотеза сдвига отклоняется на уровне достоверности α, если

$$|K| < K(\alpha),$$

где

$$K = \frac{T^*}{2}\left\{1 + \sqrt{\frac{n-1}{n-(T^*)^2}}\right\}; \quad K(\alpha) = \frac{1}{2}z_\alpha + \frac{1}{2}t_\alpha(n-1);$$

z_α — α-квантиль стандартного нормального распределения; $t_\alpha(f)$ — α-квантиль распределения Стьюдента с $f = n - 1$ степенями свободы.

Задача 235. *Имеются две выборки случайных величин*

$(n = 8)$: 1,2; 2,1; 3,8; 6,4; 7,2; 9; 11; 12,4;
$(m = 10)$: 2,1; 2,1; 6,1; 6,3; 9; 9; 11,2; 12,4; 13,6.

Необходимо проверить гипотезу сдвига критериями группы Манна–Уитни–Вилкоксона при доверительной вероятности $\alpha = 0{,}95$.

Применим критерий Манна–Уитни, для чего подсчитаем количество пар, для которых $x_i < y_j$ при всех $i = 1, \ldots, n$ и $j = 1, \ldots, m$. Например, для $i = 1$ и различных $j = 1, \ldots, 10$ имеем число таких пар, равное $\sum\limits_{j=1}^{10} h_{1j} = 10$. Далее по аналогии получаем:

i	1	2	3	4	5	6	7	8
$\sum h_{ij}$	10	8	8	6	6	4	4	2

Имеем $U = \sum_{i=1}^{8}\sum_{j=1}^{10} h_{ij} = 10+8+8+6+6+4+4+2 = 48$.

Для $\alpha = 0{,}95$ из табл. 143 находим $U_1(0,95) = 17$ и $U_2(0{,}95) = 63$.

Так как $17 < U = 48 < 63$, гипотеза сдвига отклоняется.

Для статистики Вилкоксона имеем $R = 8\cdot 10 + \dfrac{8\cdot 9}{2} - 48 = 68$.

При вычислении W-статистики необходимо иметь в виду, что у нас есть три группы совпадающих наблюдений $(2,1;2,1;2,1)$, $(9;9;9)$ и $(12,4;12,4)$, т. е. $k=3$, $t_1=3$, $t_2=3$ и $t_3=2$.

Вычисляем

$$\left\{\frac{8\cdot 10\cdot(8+10+1)}{12}\cdot\left[1-\frac{1}{(10+8)\cdot(10+8+1)\cdot(10+8-1)}\cdot\sum_{i=1}^{3} t_i\cdot\left(t_i^2-1\right)\right]\right\}^{\frac{1}{2}} = 11{,}202;$$

$$W = \frac{68-\dfrac{8\cdot 19}{2}}{11{,}202} = -0{,}714.$$

Имеем $u_{\frac{1+0,95}{2}} = u_{0,975} = 1{,}96$. Так как $|U| = 0{,}714 < 1{,}96$, гипотеза сдвига отклоняется.

Используем теперь аппроксимацию Имана. Имеем $z_{\frac{1+\alpha}{2}} = z_{0,975} = 1{,}96$.

$t_{0,975}(f = 10+8-20 = 16) = t_{0,975}(16) = 2{,}12$ (см. табл. 118). Тогда

$$J = -\frac{0,714}{2}\cdot\left[1+\sqrt{\frac{8+10-2}{8+10-1-0,714^2}}\right] = -0,708;\quad J\,(0,975) = \frac{1}{2\cdot 1,96}+\frac{1}{2\cdot 2,12} = 0,491.$$

Так как $|J| = 0{,}708 > J(0{,}975) = 0{,}491$, гипотеза сдвига не отклоняется, т. е. более точная аппроксимация отклоняет нулевую гипотезу.

Применим теперь знаковый ранговый критерий Вилкоксона (для чего ограничим в демонстрационных целях выборку y_j $m = 8$ значениями). Находим ряд разностей

$$x_i - y_i:\ -0,9;\ 0;\ -2,3;\ 0,1;\ -1,8;\ 0;\ -0,2;\ 0.$$

Ранжируем по величине значения $z_i = |x_i - y_i|$ (вверху обозначим ранг)

(ранг):	2	2	2	4	5	6	7	8
(z_i):	0	0	0	0,1	0,2	0,9	1,8	2,3

(для трех равных значений $z_i = 0$ берем средний ранг $\dfrac{1+2+3}{3} = 2$).

Величина $z_i > 0$ в ранжированном ряду имеет ранг $T = 4$. Из табл. 144 находим $T_1 = 4$ и $T_2 = 32$. Так как $T_1 = 4 \leqslant T = 4 \leqslant T_2 = 32$, гипотеза сдвига отклоняется при $\alpha = 0{,}95$.

Для T^* имеем $T^* = \dfrac{4-\dfrac{8\cdot 9}{4}}{\sqrt{8\cdot 9\cdot 17}} = -0{,}4$. Так как $|T^*| = 0{,}4 < u_{0,975} = 1{,}96$, эта аппроксимация также отклоняет гипотезу сдвига (однако следует помнить, что T-аппроксимация применима только при $n \geqslant 20$).

Более точное приближение дает (имеем в виду $t_{0,975}(f = 7) = 2{,}365$):

$$K = -\frac{0{,}4}{2}\cdot\left\{1+\sqrt{\frac{7}{8-0{,}4^2}}\right\} = -0{,}389;\quad K\,(0{,}975) = \frac{1}{2\cdot 1{,}96}+\frac{1}{2\cdot 2{,}365} = 0{,}466.$$

Так как $|K| = 0{,}389 < K(0{,}975) = 0{,}466$, гипотеза сдвига отклоняется.

4.2.1.1.2.3. Критерий Фишера–Йэйтса–Терри–Гёфдинга

Критерий рассмотрен в работах Фишера и Йэйтса [440], Терри [441] и Гёфдинга [442] и основан на статистике

$$S = \sum_{i=1}^{m} a_{m+n}(R_i)\,,$$

где $a_{m+n}(i) = M(z^i_{m+n})$ — математическое ожидание i-й порядковой статистики в выборке объема $(m+n)$ из стандартного нормального распределения; R_i — ранг значений y_i в объединенной ранжированной выборке x и y (или ранг x_i в объединенной выборке, тогда суммирование нужно вести по $i = 1, \ldots, n$).

Напомним (см. аппроксимацию 5 в разделе 1.1.1), что для $a_n(i)$ может быть использована аппроксимация

$$a_n(i) = 4{,}91\left[p^{0,14} - (1-p)^{0,14}\right], \quad \text{где} \quad p = \frac{i - \dfrac{3}{8}}{n + \dfrac{1}{4}}.$$

Гипотеза сдвига отклоняется, если $|S| < S(\alpha)$, где $S(\alpha)$ — критическая величина, некоторые значения которой приведены в табл. 145.

Таблица 145

Критические значения статистики Фишера–Йэйтса–Терри–Гёфдинга
(α — доверительная вероятность) [443]

n	m	α		n	m	α		n	m	α	
		0,90	0,95			0,90	0,95			0,90	0,95
5	3	2,12	2,74	9	4	2,62	3,04	9	7	3,15	3,68
4	4	2,27	2,59	8	5	2,77	3,23	8	8	3,17	3,73
6	3	2,33	2,69	7	6	2,83	3,29	10	7	3,22	3,78
5	4	2,42	2,72	10	4	2,67	3,11	9	8	3,27	3,82
7	3	2,32	2,66	9	5	2,82	3,82	10	8	3,35	3,94
6	4	2,46	2,82	8	6	2,93	3,40	9	9	3,38	3,96
5	5	2,58	2,92	7	7	2,95	3,44	11	8	3,42	4,02
7	4	2,47	2,92	10	5	2,89	3,40	10	9	3,46	4,07
6	5	2,60	3,00	9	6	3,00	3,52	12	8	3,50	4,11
8	4	2,55	3,00	8	7	3,06	3,58	11	9	3,55	4,17
7	5	2,66	3,10	11	5	2,94	3,45	10	10	3,57	4,20
6	6	2,75	3,19	10	6	3,07	3,59				

При $m, n \geqslant 10$ распределение S удовлетворительно аппроксимируется нормальным со средним $\mu = \mathbf{M}(S) = 0$ и дисперсией

$$\mathbf{D}(S) = \frac{mn}{(m+n)(m+n-1)} \sum_{i=1}^{m+n} a^2_{m+n}(i)\,.$$

В этом случае гипотеза сдвига отклоняется, если

$$|S^*| = \left|\frac{S}{\mathbf{D}\,(S)}\right| < u_{\frac{1+\alpha}{2}}\,,$$

где u_γ — γ-квантиль стандартного нормального распределения.

Задача 236. *Для двух выборок*

$$(n=8)\quad x_i\colon\ 1,\ 3,\ 4,\ 7,\ 9,\ 10,\ 15,\ 16;$$
$$(m=7)\quad y_i\colon\ 6,\ 8,\ 11,\ 14,\ 18,\ 21,\ 26$$

проверить гипотезу сдвига критерием Фишера–Йэйтса–Терри–Гёфдинга при доверительной вероятности $\alpha = 0{,}95$.

Ранжируем совместную выборку, отмечая в ней ранги величин x и y (вверху отмечен ранг, внизу принадлежность к выборке x или y)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	3	4	6	7	8	9	10	11	14	15	16	18	21	26
x	x	x	y	x	y	x	x	y	y	x	x	y	y	y

Следовательно, ранги членов выборки x в общем ряду будут

$$R_i\colon 1,\ 2,\ 3,\ 5,\ 7,\ 8,\ 11,\ 12.$$

Для них находим $a_{15}(i) = 4{,}91\cdot\left[p^{0,14} - (1-p)^{0,14}\right]$, где $p = \dfrac{i-\frac{3}{8}}{15+\frac{1}{4}}$.

В результате имеем:

i	1	2	3	5	7	8	11	12
$a_{15}(i)$	−1,7419	−1,2444	−0,94388	−0,51307	−0,16441	0	0,51307	0,71156

Далее вычисляем

$$S = \sum_{i=1}^{8} a_{15}(R_i) = -1{,}7419 - 1{,}2444 - \ldots + 0{,}51307 + 0{,}71156 = -3{,}383.$$

Из табл. 145 для $\alpha = 0{,}90$, $n = 8$ и $m = 7$ имеем $S(0{,}95) = 3{,}58$.
Так как $|S| = 3{,}383 < S(0{,}95) = 3{,}58$, гипотеза сдвига отклоняется.
Для нормального приближения вычисляем дополнительно:

i	4	6	9	10	13	14	15
$a_{15}(i)$	−0,71156	−0,33358	0,16441	0,33349	0,94388	1,2444	1,7419

Имеем

$$\sum_{i=1}^{15} a_{15}^2(i) = 12{,}763;\qquad \mathbf{D}(S) = \frac{8\cdot 7}{(8+7)\cdot(8+7-1)} = \frac{3{,}38}{1{,}845} = 1{,}832.$$

Так как $|S^*| = 1{,}832 < u_{0,975} = 1{,}96$, гипотеза сдвига отклоняется.

4.2.1.1.2.4. Критерий Ван дер Вардена [2]

Статистика критерия имеет вид

$$X = \sum_{i=1}^{m} u_{\frac{R_i}{m+n+1}},$$

где u_γ — γ-квантиль стандартного нормального распределения.

Для вычисления квантилей $u_{\frac{R_i}{m+n+1}}$ может быть применено приближение (см. аппроксимацию 15 в разделе 1.1.1)

$$u_{\frac{R_i}{n+m+1}} \approx 4{,}91\left[\left(\frac{R_i}{n+m+1}\right)^{0,14} - \left(1-\frac{R_i}{n+m+1}\right)^{0,14}\right].$$

Гипотеза сдвига отклоняется, если $|X| < x_\alpha$, где x_α — критическое значение, приведенное в табл. 146.

Таблица 146

Критические значения X_α статистики Ван дер Вардена
(α — доверительная вероятность) [2, 443]

n	m	α		n	m	α		n	m	α	
		0,90	0,95			0,90	0,95			0,90	0,95
4	3	1,82	2,14	7	6	2,56	2,98	9	8	2,99	3,51
4	4	1,98	2,27	9	5	2,57	3,04	12	6	2,91	3,43
5	3	1,84	2,41	8	6	2,65	3,07	11	7	3,02	3,54
6	3	2,05	2,37	7	7	2,69	3,12	10	8	3,08	3,61
5	4	2,12	2,39	10	5	2,61	3,07	9	9	3,09	3,63
6	4	2,18	2,50	9	6	2,71	3,20	13	6	2,97	3,49
5	5	2,29	2,60	8	7	2,76	3,25	12	7	3,08	3,62
7	4	2,26	2,59	11	5	2,68	3,14	11	8	3,15	3,70
6	5	2,32	2,69	10	6	2,80	3,28	10	9	3,19	3,75
8	4	2,32	2,69	9	7	2,87	3,36	14	6	3,02	3,56
7	5	2,40	2,79	8	8	2,90	3,40	13	7	3,15	3,70
6	6	2,45	2,85	12	5	2,73	3,20	12	8	3,23	3,80
9	4	2,36	2,76	11	6	2,86	3,36	11	9	3,28	3,86
8	5	2,50	2,92	10	7	2,95	3,46	10	10	3,30	3,88

При $n, m \geqslant 20$ распределение X удовлетворительно описывается нормальным со средним $\mu = \mathbf{M}(x) = 0$ и дисперсией

$$\mathbf{D}(X) = \frac{mn}{(m+n)(m+n-1)} \sum_{i=1}^{m+n} u^2_{\frac{R_i}{m+n+1}}.$$

Если $|X^*| = \left|\dfrac{X}{\mathbf{D}(X)}\right| < u_{\frac{1+\alpha}{2}}$, гипотеза сдвига отклоняется с достоверностью α.

Вспомогательные величины

$$K = \frac{1}{m+n} \sum_{i=1}^{m+n} u^2_{\frac{R_i}{n+m+1}}$$

приведены в табл. 147.

При $n + m \to \infty$ эффективность критерия Ван дер Вардена не уступает эффективности критерия Стьюдента.

Задача 237. *Проверить гипотезу сдвига в условиях задачи 236 критерием Ван дер Вардена.*

Для рангов выборки x в общем ранжированном ряду

$$R_i\colon\ 1,\ 2,\ 3,\ 5,\ 7,\ 8,\ 11,\ 12$$

находим

$$u_{\frac{1}{8+7+1}} = u_{\frac{1}{16}} = 4{,}91 \cdot \left[\left(\frac{1}{16}\right)^{0,14} - \left(\frac{15}{16}\right)^{0,14}\right] = -1{,}5354;$$

$$u_{\frac{2}{16}} = -1{,}1492; \quad u_{\frac{3}{16}} = -0{,}8851; \quad u_{\frac{5}{16}} = -0{,}4869; \quad u_{\frac{7}{16}} = -0{,}1566;$$

$$u_{\frac{8}{16}} = 0; \quad u_{\frac{11}{16}} = 0{,}4869; \quad u_{\frac{12}{16}} = 0{,}6723.$$

Имеем $X = \sum_{i=1}^{8} u_{R_{\frac{i}{m+n+1}}} = -3{,}054$. Из табл. 146 для $n = 8$, $m = 7$ находим $x_{0,95} = 3{,}25$. Так как $|X| = 3{,}054 < x_{0,95} = 3{,}25$, гипотеза сдвига отклоняется.

Используем теперь нормальное приближение. Из табл. 147 находим

$$K = \frac{1}{m+n} \cdot \sum_{i=1}^{m+n} u^2_{\frac{R_i}{m+n+1}} = 0{,}705; \quad \mathbf{D}\,(X) = \frac{m \cdot n}{m+n-1} \cdot K = \frac{7 \cdot 8}{14} \cdot 0{,}705 = 2{,}82;$$

$$\sqrt{\mathbf{D}\,(X)} = 1{,}679.$$

Окончательно $|X^*| = \dfrac{3{,}054}{1{,}679} = 1{,}819$.

Так как $|X^*| = 1{,}819 < u_{0,95} = 1{,}96$, нулевая гипотеза сдвига отклоняется.

Таблица 147

Вспомогательные величины K критерия Ван дер Вардена [2]

$m+n$	K	$m+n$	K	$m+n$	K	$m+n$	K	$m+n$	K
5	0,449	21	0,763	37	0,839	53	0,876	78	0,907
6	0,497	22	0,770	38	0,842	54	0,877	80	0,908
7	0,537	23	0,777	39	0,845	55	0,879	82	0,910
8	0,570	24	0,783	40	0,848	56	0,880	84	0,912
9	0,598	25	0,789	41	0,850	57	0,882	86	0,913
10	0,622	26	0,794	42	0,853	58	0,884	88	0,915
11	0,642	27	0,799	43	0,855	59	0,885	90	0,916
12	0,661	28	0,804	44	0,858	60	0,887	92	0,918
13	0,667	29	0,809	45	0,860	62	0,889	94	0,919
14	0,692	30	0,813	46	0,862	64	0,892	96	0,920
15	0,705	31	0,817	47	0,864	66	0,894	98	0,922
16	0,716	32	0,821	48	0,866	68	0,897	100	0,923
17	0,727	33	0,825	49	0,868	70	0,899	110	0,928
18	0,737	34	0,829	50	0,870	72	0,901	120	0,933
19	0,746	35	0,833	51	0,872	74	0,903	130	0,937
20	0,755	36	0,836	52	0,874	76	0,905	140	0,940

4.2.1.1.2.5. Медианный критерий

Статистика критерия строится следующим образом [365]. Находится медиана $\mathbf{Me}(x, y)$ общего упорядоченного ряда (x, y), и подсчитывается число наблюдений выборки x, превосходящих медиану (если $(m+n)$ нечетно и медиана принадлежит выборке x, то это число увеличивается на $1/2$). Тогда статистика критерия может быть записана как

$$S = \sum_{i=1}^{m} \frac{1}{2}\left[\delta\left(R_i - \frac{m+n+1}{2}\right) + 1\right], \quad \text{где} \quad \delta(x) = \begin{cases} 1, & x > 0; \\ 0, & x = 0; \\ -1, & x < 0. \end{cases}$$

При $n, m \geqslant 10$ распределение S удовлетворительно описывается нормальным [365] со средним $\mu = \mathbf{M}(S) = \dfrac{m}{2}$ и дисперсией

$$\mathbf{D}(S) = \begin{cases} \dfrac{mn}{4(m+n-1)}, & \text{если } m+n = 2k; \\ \dfrac{mn}{4(m+n)}, & \text{если } m+n = 2k-1. \end{cases}$$

Если

$$|S^*| = \frac{|S - \mathbf{M}(S)|}{\sqrt{\mathbf{D}(S)}} < U_{\frac{1+\alpha}{2}},$$

то с достоверностью α гипотеза сдвига отклоняется.

Иногда применяется другая форма медианного критерия [444]. Пусть A и C — количества элементов выборки x, соответственно бо́льших и меньших медианы объединенной выборки, а B и D — аналогичные числа для выборки y. Тогда статистикой критерия сдвига является величина

$$\chi^2 = \frac{(n+m)\left(|AD - BC| - \dfrac{n+m}{2}\right)^2}{(A+B)(C+D) + (A+C)(B+D)},$$

имеющая, при отсутствии сдвига, распределение хи-квадрат с $f = 1$ степенью свободы.

Критерий неприменим, если A, B, C или $D < 5$ и $n + m < 40$. Для этих условий следует пользоваться точным критерием, критические значения которого D и C для заданных A и B приведены в [44] (из-за громоздкости таблиц мы их здесь не приводим).

Эффективность медианного критерия по сравнению с критерием Стьюдента в случае нормального распределения равна $2/\pi \approx 0{,}64$.

Задача 238. *Даны две выборки случайных величин:*

$(m = 20)$ x: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20;
$(n = 21)$ y: 9 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49.

Необходимо проверить гипотезу сдвига медианным критерием при доверительной вероятности $\alpha = 0{,}95$.

Ранжируем совместный ряд (x, y) (вверху обозначен ранг члена выборки в общем ряду, для совпадающих значений ранги усреднены)

1	2	3	4	5	6	7	8	9,5	9,5	11	12,5	12,5	14	15,5	15,5	17	18,5	18,5	20
1	2	3	4	5	6	7	8	9	9	10	11	11	12	13	13	14	15	15	16
x	x	x	x	x	x	x	x	x	y	x	x	y	x	x	y	x	x	y	x

21,5	21,5	23	24,5	24,5	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
17	17	18	19	19	20	21	23	25	27	29	31	33	35	37	39	41	43	45	47	49
x	y	x	x	y	x	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y

Легко видеть, что медианой этого ряда является член выборки с порядковым номером $\dfrac{m+n+1}{2} = \dfrac{20+21+1}{2} = 21$, это $\tilde{x} = 17$.

Вычисляем

$$S = \sum_{i=1}^{20} \frac{1}{2} \cdot \left[\delta\left(R_i - \frac{m+n+1}{2}\right) + 1\right] = \sum_{i=1}^{20} \frac{1}{2} \cdot \delta[(R_i - 21) + 1] = \frac{1}{2} \times$$

$$\times(-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1+0+1+1+1) + \frac{20}{2} = 3{,}5;$$

$$\mathbf{M}(S) = \frac{m}{2} = \frac{20}{2} = 10; \quad \mathbf{D}(S) = \frac{m \cdot n}{4 \cdot (m+n)} = \frac{20 \cdot 21}{4 \cdot 41} = 2{,}561 \quad (\sqrt{\mathbf{D}(S)} = 1{,}60).$$

Тогда $|S^*| = \dfrac{|3{,}5 - 10|}{1{,}60} = 4{,}06$. Так как $|S^*| = 4{,}06 > u_{\frac{1+0{,}95}{2}} = 1{,}96$, гипотеза сдвига не отклоняется.

Рассмотрим вторую модификацию критерия. В нашем случае $A = 3$, $C = B = 16$ и $D = 4$. Тогда

$$\chi^2 = \frac{(21+20)\cdot\left(|3\cdot 4 - 16\cdot 16| - \dfrac{21+20}{2}\right)^2}{(3+16)\cdot(16+4)+(3+16)\cdot(16+4)} = 2694{,}8.$$

что, конечно же, чудовищно превышает критическое значение $\chi^2_{0,95}(1) = 3{,}84$. Следует помнить, что критерий применим только при $n + m < 40$ и $A, B, C, D < 5$. Мы его применили некорректно, однако основной цели достигли — продемонстрировали технику применения критерия.

4.2.1.1.2.6. Критерий Мостеллера

Рассмотрен в [445] и формулируется следующим образом. Гипотеза равенства средних двух выборок одинакового объема $n > 5$ отклоняется с доверительной вероятностью 0,95, если 5 (при $n \leqslant 25$) или 6 (при $n > 25$) наибольших или наименьших значений содержатся в одной и той же выборке. Критерий имеет низкую мощность и может быть рекомендован только для быстрой грубой проверки гипотез сдвига.

Задача 239. *В условиях задачи* 238 (*исключив значение* $y = 49$ *для того, чтобы выровнять объемы выборок* $n = m = 20$) *проверить гипотезу сдвига критерием Мостеллера.*

Видим, что здесь 14 наибольших значений принадлежат выборке y, что намного превышает критическое значение, равное 5. Следовательно, можно уверенно признать наличие сдвига в параметре положения двух выборок.

4.2.1.1.2.7. Критерий Розенбаума

Предложен в [446] для двух выборок равного объема. Если не менее 5 (для $n \geqslant 16$ и $\alpha = 0{,}95$) или 7 (для $n \geqslant 20$ и $\alpha = 0{,}99$) значений одной выборки находятся вне размаха второй выборки, то нулевая гипотеза отсутствия сдвига на указанных уровнях достоверности отклоняется.

Критерий рекомендуется использовать для быстрой приближенной проверки гипотезы сдвига.

Задача 240. *Проверить гипотезу сдвига в условиях задачи* 238 *критерием Розенбаума.*

Вне размаха выборки y (от 9 до 47) находятся 8 значений выборки x, что больше критического числа 5 и позволяет предположить наличие сдвига в параметрах положения двух выборок.

4.2.1.1.2.8. Критерий Хаги

Рассмотрен в [447]. Пусть A и B' — количества наблюдений среди $x_1, \ldots, x_n$, бо́льших, чем $\max\limits_{1\leqslant j\leqslant n} y_j$, и меньших, чем $\min\limits_{1\leqslant j\leqslant n} y_j$ соответственно, и пусть A' и B — количества наблюдений среди $y_1, \ldots, y_n$, бо́льших, чем $\max\limits_{1\leqslant i\leqslant m} x_i$, и меньших, чем $\min\limits_{1\leqslant i\leqslant m} x_i$ соответственно. Критерий Хаги основан на статистике

$$T = A + B - A' - B'.$$

Гипотеза сдвига принимается на уровне значимости α, если $|T| > T_\alpha$, где T_α — критическое значение, приведенное в табл. 148.

При $n, m > 25$ применимо соотношение [448]

$$\mathbf{P}(T \geqslant k) = \frac{(p^2+q)p}{(p-q)(1+p)}p^k - \frac{(p^2+q)q}{(p-q)(1+q)},$$

Таблица 148

Критические значения T_α критерия Хаги (α — уровень значимости) [448]

$n=m$	α		$n=m$	α		$n=m$	α		$n=m$	α	
	0,05	0,10		0,05	0,10		0,05	0,10		0,05	0,10
4	6	5	10	7	6	16	7	6	22	7	6
5	6	5	11	7	6	17	7	6	23	7	6
6	6	5	12	7	6	18	7	6	24	7	6
7	6	5	13	7	6	19	7	6			
8	7	6	14	7	6	20	7	6			
9	7	6	15	7	6	21	7	6			

где

$$p=\frac{m}{n+m};\quad q=1-p;\quad p\neq\frac{1}{2}\quad \left(\text{при } p=\frac{1}{2}\colon \mathbf{P}(T\geqslant k)=\left(\frac{k}{2}+\frac{1}{3}\right)2^{-k}\right).$$

Гипотеза сдвига принимается, если вычисленное значение $\mathbf{P}(T\geqslant k)<\frac{\alpha}{2}$, где α — заданный уровень значимости.

Задача 241. *Для данных задачи 238 проверить гипотезу сдвига критерием Хаги на уровне значимости $\alpha=0{,}05$ (принимаем $n=m=20$).*

Имеем $A=0$, $B'=8$, $A'=14$, $B=0$ и $T=0+0-14-8=-22$.
Из табл. 148 для $n=m=20$ находим критическое значение $T_{0{,}05}=7$.
Так как $|T|=22>T_{0{,}05}=7$, гипотеза сдвига уверенно принимается.
Находим далее

$$\mathbf{P}(T\geqslant 22)=\left(\frac{22}{2}+\frac{1}{3}\right)\cdot 2^{-22}=2{,}7\cdot 10^{-6}.$$

Легко видеть, что $\mathbf{P}(T\geqslant 22)\ll 0{,}05/2=0{,}025$, что также позволяет принять гипотезу сдвига.

4.2.1.1.2.9. *E*-критерий

Предложен Гаеком и Шидаком [365, 449]. В обозначениях, введенных для критерия Хага (см. раздел 4.2.1.1.2.8), статистика E-критерия записывается как

$$E=\left|\min(A,B)-\min(A',B')\right|.$$

Таблицы критических значений E-критерия приведены в [450]. Однако мы рекомендуем вместо таблиц использовать достаточно простую формулу

$$\mathbf{P}(E\geqslant k)=\frac{C_{m+n-2k}^{m-k}}{C_{m+n}^{n}},\quad \text{где}\quad k=1,2,\ldots,\min(m,n).$$

Так как $\mathbf{P}(E\geqslant k)=\mathbf{P}(E\leqslant k)$, то на уровне значимости α гипотеза сдвига будет принята, если $\mathbf{P}(E\geqslant k)$ или $\mathbf{P}(E\leqslant k)$ будет меньше, чем $\alpha/2$.

Задача 242. *Проверить E-критерием гипотезу сдвига в условиях задачи 238.*

Имеем $A=0$, $B'=8$, $A'=14$ и $B=0$.
Тогда $E=|\min(0{,}0)-\min(8{,}14)|=8$ и

$$\mathbf{P}(E\geqslant 8)=\frac{C_{20+20-2\cdot 8}^{20-8}}{C_{20+20}^{20}}=\frac{C_{14}^{2}}{C_{40}^{20}}=1{,}96\cdot 10^{-5}.$$

Легко видеть, что $\mathbf{P}(E\geqslant 8)\ll\alpha/2=0{,}025$, что позволяет уверенно принять гипотезу сдвига, так как наблюдаемое значение $E=8$ очень маловероятно при отсутствии сдвига в параметрах положения двух выборок.

4.2.1.2. Сравнение параметров сдвига нескольких ($k > 2$) совокупностей

4.2.1.2.1. Критерий Крускала–Уоллиса

Рассмотрен в [451]. Пусть в нашем распоряжении имеются k выборок случайных величин

$$x_{11}, x_{12}, \ldots, x_{1n_1}; \quad x_{21}, x_{22}, \ldots, x_{2n_2}; \quad \ldots \quad ; x_{k1}, x_{k2}, \ldots, x_{kn_k}.$$

Упорядочим все $N = \sum\limits_{i=1}^{k} n_i$ элементов выборок по возрастанию и обозначим через R_{ij} ранг j-го элемента i-й выборки в общем упорядоченном ряду.

Статистика критерия Крускала–Уоллиса для проверки гипотезы о наличии сдвига в параметрах положения двух сравниваемых выборок имеет вид

$$H = \sum_{i=1}^{k} \left(1 - \frac{n_i}{N}\right) \left\{ \frac{\bar{R}_i - \dfrac{N+1}{2}}{\sqrt{\dfrac{(N-n_i)(N+1)}{12n_i}}} \right\}^{\frac{1}{2}} =$$
$$= \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(\bar{R}_i - \frac{N+1}{2}\right)^2 = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1),$$

где $R_i = \sum\limits_{j=1}^{N} R_{ij}$; $\bar{R}_i = \dfrac{1}{n_i} R_i$.

Легко видеть, что критерий Крускала–Уоллиса является многомерным обобщением двухвыборочного критерия Вилкоксона–Манна–Уитни (см раздел 4.2.1.1.2.2).

Гипотеза сдвига отклоняется на уровне значимости α, если $H \geqslant H_\alpha$, где H_α — критическое значение, приведенное в табл. 149 для $k \leqslant 5$ и $n_i \leqslant 8$. При $n_i \geqslant 5$ применимы различные аппроксимации. Укажем некоторые из них.

Таблица 149

Критические значения критерия Крускала–Уоллиса
(α — уровень значимости) [18, 57, 444]

n_1	n_2	n_3	α = 0,1	α = 0,05	n_1	n_2	n_3	α = 0,1	α = 0,05	n_1	n_2	n_3	α = 0,1	α = 0,05
$k = 3$														
2	2	2	4,571		5	4	2	4,541	5,273	6	5	2	4,596	5,338
3	2	2	4,500	4,714	5	4	3	4,549	5,656	6	5	3	4,535	5,602
3	3	2	4,556	5,361	5	4	4	4,668	5,657	6	5	4	4,522	5,661
3	3	3	4,622	5,600	5	5	2	4,623	5,338	6	5	5	4,547	5,729
4	2	2	4,458	5,333	5	5	3	4,545	5,705	6	6	2	4,438	5,410
4	3	2	4,511	5,444	5	5	4	4,523	5,666	6	6	3	4,558	5,625
4	3	3	4,709	5,791	5	5	5	4,560	5,780	6	6	4	4,548	5,724
4	4	2	4,555	5,455	6	2	2	4,545	5,345	6	6	5	4,542	5,765
4	4	3	4,545	5,598	6	3	2	4,682	5,348	6	6	6	4,643	5,801
4	4	4	4,654	5,692	6	3	3	4,590	5,615	7	7	7	4,594	5,819
5	2	2	4,373	5,160	6	4	2	4,494	5,340	8	8	8	4,595	5,805
5	3	2	4,651	5,251	6	4	3	4,604	5,610					
5	3	3	4,533	5,648	6	4	4	4,595	5,681					

Окончание таблицы 149

$k=4$											
n_1	n_2	n_3	n_4	α		n_1	n_2	n_3	n_4	α	
				0,1	0,05					0,1	0,05
2	2	2	2	5,667	6,167	4	3	2	2	5,750	6,621
3	2	2	2	5,664	6,333	4	3	3	2	5,872	6,795
3	3	2	2	5,745	6,527	4	3	3	3	6,016	6,984
3	3	3	2	5,879	6,727	4	4	2	2	5,808	6,731
3	3	3	3	6,026	7,000	4	4	2	2	5,901	6,874
4	2	2	2	5,755	6,545						

$k=5$													
n_1	n_2	n_3	n_4	n_5	α		n_1	n_2	n_3	n_4	n_5	α	
					0,1	0,05						0,1	0,05
4	4	4	2		5,914	6,957	3	3	2	2	2	7,026	7,910
4	4	4	3		6,042	7,142	3	3	3	2	2	7,121	8,044
4	4	4	4		6,088	7,235	3	3	3	3	2	7,210	8,200
2	2	2	2	2	6,982	7,418	3	3	3	3	3	7,333	8,333
3	2	2	2	2	6,955	7,682							

Аппроксимация Крускала–Уоллиса [451]
Пусть

$$M = \frac{N^3 - \sum_{i=1}^{k} n_i^3}{N(N+1)}; \quad \nu_1 = (k-1)\frac{(k-1)(M-k+1)-V}{\frac{1}{2}MV}; \quad \nu_2 = \frac{M-k+1}{k-1}\nu_1;$$

$$V = 2(k-1) - \frac{2\{3k^2 - 6k + N(2k^2 - 6k + 1)\}}{5N(N+1)} - \frac{6}{5}\sum_{i=1}^{k}\frac{1}{n_i}.$$

Тогда статистика

$$F = \frac{H(M-k+1)}{(k-1)(M-H)}$$

будет иметь при отсутствии сдвига F-распределение с ν_1 и ν_2 степенями свободы. Таким образом, нулевая гипотеза отклоняется с достоверностью α, если $F > F_\alpha(\nu_1, \nu_2)$.

Аппроксимация Имана–Давенпорта [452]

В соответствии с ней нулевая гипотеза сдвига отклоняется с достоверностью α, если $J \geqslant J_\alpha$, где

$$J = \frac{H}{2}\left(1 + \frac{N-k}{N-1-H}\right); \quad J_\alpha = \left\{(k-1)\,F_\alpha(k-1; N-k) + \chi^2_\alpha(k-1)\right\},$$

$F_\alpha(f_1, f_2)$ и $\chi^2_\alpha(a)$ — соответственно критические значения статистик Фишера и хи-квадрат с соответствующими степенями свободы.

Это более точная аппроксимация, чем аппроксимация Крускала–Уоллиса. При наличии связанных рангов (т. е. когда совпадают значения величин из разных

выборок и им присваивается одинаковые средние ранги) необходимо использовать модифицированную статистику $H^* = H\left\{1 - \left(\sum_{j=1}^{q} \frac{T_j}{N^3 - N}\right)\right\}^{-1}$, где $T_j = t_j^3 - t_j$; t_j — размер j-й группы одинаковых элементов; q — количество групп одинаковых элементов.

При $n_i \geqslant 20$ справедлива аппроксимация распределения статистики H χ^2-распределением с $f = k - 1$ степенями свободы, т. е. нулевая гипотеза отклоняется, если $H \geqslant \chi^2_\alpha(k-1)$.

Задача 243. *В результате наблюдений получены пять выборок случайных величин* ($k = 5$)

x_{1j}: 1, 2, 3, 4, 5, 6 ($n_1 = 6$); x_{2j}: 3, 4, 5, 6, 7 ($n_2 = 5$); x_{3j}: 7, 8, 9 ($n_3 = 3$); x_{4j}: 1, 5, 7, 8, 10, 12 ($n_4 = 6$); x_{5j}: 10, 11, 13, 14, 16, 18, 20 ($n_5 = 7$).

Необходимо проверить гипотезу об отсутствии сдвига между параметрами положения в выборках критерием Крускала–Уоллиса на уровне значимости $\alpha = 0{,}05$.

Ранжируем совместно все $N = \sum_{i=1}^{5} n_i = 6 + 5 + 5 + 6 + 7 = 27$ выборочных значений x_{ij} (i — номер выборки; R_{ij} — ранг j-го наблюдения в i-й выборке). Результаты сведем в таблицу, усредняя ранги совпадающих наблюдений:

i	x_{ij}	R_i	i	x_{ij}	R_i	i	x_{ij}	R_i	i	x_{ij}	R_i
1	1	1,5	1	5	9	4	7	14	4	12	22
4	1	1,5	2	5	9	3	8	16,5	5	13	23
1	2	3	4	5	9	4	8	16,5	5	14	24
1	3	4,5	1	6	11,5	3	9	18	5	16	25
2	3	4,5	2	6	11,5	4	10	19,5	5	18	26
1	4	6,5	2	7	14	5	10	19,5	5	20	27
2	4	6,5	3	7	14	5	11	21			

Далее подсчитываем

$$R_1 = \sum_{j+1}^{27} R_{1j} = 1{,}5 + 3 + 4{,}5 + 6{,}5 + 9 + 11{,}5 = 36 \ (\bar{R}_1 = \frac{36}{6} = 6);$$

$$R_2 = 4{,}5 + 16{,}5 + 9 + 11{,}5 + 14 = 45{,}5 \ (\bar{R}_2 = \frac{45{,}5}{5} = 9{,}1);$$

$$R_3 = 48{,}5 \ (\bar{R}_3 = 16{,}166); \quad R_4 = 82{,}5 \ (\bar{R}_4 = 13{,}75); \quad R_5 = 165{,}5 \ (\bar{R}_5 = 23{,}643).$$

Тогда

$$H = \frac{12}{27 \cdot 28} \cdot \sum_{i=1}^{5} \frac{R_i^2}{n_i} - 3 \cdot 28 = 18{,}562.$$

Используя аппроксимацию Крускала–Уоллиса, имеем

$$M = \frac{27^3 - \sum_{i=1}^{3} n_i^3}{27 \cdot 28} = 24{,}8095; \quad \nu_1 = 3{,}8274; \quad \nu_2 = 19{,}91;$$

$$V = 2 \cdot 4 - \frac{2 \cdot \{3 \cdot 5^2 - 6 \cdot 5 + 27 \cdot (2 \cdot 25 - 6 \cdot 5 + 1)\}}{5 \cdot 27 \cdot 28} - \frac{6}{5} \cdot \sum_{i=1}^{5} \frac{1}{n_i} = 6{,}464.$$

Учитывая, что у нас были группы совпадающих рангов (всего было $q = 8$ совпадающих групп с $t_1 = 2$, $t_2 = 2$, $t_3 = 2$, $t_4 = 3$, $t_5 = 2$, $t_6 = 3$, $t_7 = 2$, $t_8 = 2$), находим

$$H^* = \frac{H}{1 - \frac{1}{27^3 - 27} \cdot \left[(2^3 - 2) \cdot 6 + 2 \cdot (3^3 - 3)\right]} = 18{,}642;$$

$$F = \frac{18{,}642 \cdot (24{,}8095 - 5 + 1)}{4 \cdot (24{,}8095 - 18{,}642)} = 15{,}725.$$

Из таблиц F-распределения (или аппроксимаций — см. раздел 1.1.10) находим для уровня достоверности $1 - \alpha = 0{,}95$: $F_{0,95}(\nu_1 = 3{,}83; \nu_2 = 19{,}91) = 3{,}05$.

Так как $F = 15{,}725 > F_{0,95}(3{,}83; 19{,}91) = 3{,}05$, гипотеза сдвига не отклоняется (т. е. сдвиг признается значимым).

Используем теперь аппроксимацию Имана–Давенпорта:

$$J = \frac{18{,}642}{2} \cdot \left(1 + \frac{27 - 5}{27 - 1 - 18{,}642}\right) = 37{,}19; \quad J_{0,95} = \frac{1}{2} \cdot \left[4 \cdot F_{0,95}(4; 22) + \chi^2_{0,95}(4)\right].$$

Из таблиц находим $F_{0,95}(4; 22) = 2{,}82$ и $\chi^2_{0,95}(4) = 9{,}49$ (аппроксимации см. в разделах 1.1.8, 1.1.10). Тогда $J_{0,95} = \frac{1}{2} \cdot (4 \cdot 2{,}82 + 9{,}49) = 10{,}385$.

Так как $J = 37{,}19 > J_{0,95} = 10{,}385$, гипотеза сдвига не отклоняется.

4.2.1.2.2. Критерий Неменьи

Критерий применим для выборок равного объема n (т. е. $N = nk$). Статистика критерия, предложенного Неменьи [453], в обозначениях, принятых для критерия Крускала–Уоллиса (см. раздел 4.2.1.2.1), имеет вид $D_{l,m} = |R_l - R_m|$, ($l \neq m$; $l, m = 1, 2, \ldots, k$).

Гипотеза сдвига считается принятой, если $D_{l,m} \geqslant D_{l,m}(\alpha)$, где $D_{l,m}(\alpha)$ — критические значения, приведенные в табл. 150.

Задача 244. *Для $k = 4$ выборок равного объема $n = 5$*

$$x_{1j}:\ 1,\ 2,\ 3,\ 4,\ 5; \qquad x_{2j}:\ 3,\ 4,\ 5,\ 6,\ 7;$$
$$x_{3j}:\ 4,\ 5,\ 6,\ 7,\ 8; \qquad x_{4j}:\ 7,\ 8,\ 9,\ 10,\ 11$$

проверить гипотезу сдвига критерием Неменьи при достоверности $\alpha = 0{,}95$.

Упорядочиваем по возрастанию ряд $N = n \cdot k = 5 \cdot 4 = 20$ значений x_{ij} (i — номер выборки). Результаты сводим в таблицу, ранги совпадающих наблюдений усредняем:

i	x_{ij}	R_{ij}	i	x_{ij}	R_{ij}	i	x_{ij}	R_{ij}	i	x_{ij}	R_{ij}
1	1	1	2	4	6	2	6	11,5	3	8	16,5
1	2	2	3	4	6	3	6	11,5	4	8	16,5
1	3	3,5	1	5	9	2	7	14	4	9	18
2	3	3,5	2	5	9	3	7	14	4	10	19
1	4	6	3	5	9	4	7	14	4	11	20

Вычисляем

$$R_1 = \sum_{j=1}^{20} R_{1j} = 1 + 2 + 3{,}5 + 6 + 9 = 21{,}5; \quad R_2 = \sum_{j=1}^{20} R_{2j} = 3{,}5 + 6 + 9 + 11{,}5 + 14 = 44;$$

$$R_3 = \sum_{j=1}^{20} R_{3j} = 6 + 9 + 11{,}5 + 14 + 16{,}5 = 57;$$

$$R_4 = \sum_{j=1}^{20} R_{4j} = 14 + 16{,}5 + 18 + 19 + 20 = 87{,}5.$$

Таблица 150

Критические значения $D_{l,m}(\alpha)$ критерия Неменьи [9]

n	k							
	3	4	5	6	7	8	9	10
Доверительная вероятность $\alpha = 0{,}90$								
3	13,8	20,2	26,9	33,9	40,9	48,1	55,5	63,0
4	20,9	30,9	41,2	51,8	62,6	73,8	85,1	96,5
5	29,0	42,9	57,2	72,1	87,3	102,8	118,6	134,6
6	37,9	56,1	75,0	94,5	114,4	134,8	155,6	222,3
7	47,6	70,5	94,3	118,8	144,0	169,6	195,8	271,4
8	58,0	86,0	115,0	145,0	175,7	207,0	239,0	323,6
9	69,1	102,4	137,0	172,8	209,4	246,8	284,9	378,8
10	80,8	119,8	160,3	202,2	245,1	288,9	333,5	436,8
11	93,1	138,0	184,8	233,1	282,6	333,1	384,6	497,5
12	105,9	157,1	210,4	265,4	321,8	379,3	438,0	560,8
13	119,3	177,0	237,1	299,1	362,7	427,6	493,7	626,6
14	133,2	197,7	264,8	334,1	405,1	477,7	551,6	694,8
15	147,6	219,1	293,0	370,4	449,2	529,6	611,6	765,2
16	162,5	241,3	323,3	407,9	494,7	583,3	673,6	837,7
17	177,9	264,2	353,9	446,6	541,6	638,7	737,6	912,8
18	193,7	287,7	385,5	486,5	590,0	695,7	803,4	989,7
19	210,0	311,9	417,9	527,5	639,7	754,3	871,2	1068,8
20	226,7	336,7	451,2	568,5	690,7	814,5	940,7	1149,8
21	243,8	362,2	485,4	612,6	743,0	876,2	1012,0	1232,7
22	261,3	388,2	520,4	656,8	796,6	939,4	1085,0	1317,6
23	279,2	414,9	556,1	702,0	851,4	1001,1	1159,7	1404,3
24	297,5	442,2	592,7	748,1	907,4	1070,2	1236,0	1492,9
25	316,2	470,0	630,0	795,3	964,6	1137,6	1314,0	1591,0
Доверительная вероятность $\alpha = 0{,}95$								
3	15,7	22,7	29,9	37,3	44,8	52,5	60,3	68,2
4	23,9	34,6	45,6	57,0	60,6	80,4	92,4	104,6
5	33,1	48,1	63,5	79,3	95,5	112,0	128,8	145,8
6	43,3	62,9	83,2	104,0	125,3	147,0	169,1	191,4
7	54,4	79,1	104,6	130,8	157,6	184,9	212,8	240,9
8	66,2	96,4	127,6	159,6	192,4	225,7	259,7	294,1
9	78,9	114,8	152,0	190,2	229,3	269,1	309,6	350,6
10	92,3	134,3	177,8	222,6	268,4	315,0	362,4	410,5
11	106,3	154,8	205,0	256,6	309,4	363,2	417,9	473,3
12	120,9	176,2	233,4	292,2	352,4	413,6	476,0	539,1
13	136,2	198,5	263,0	329,3	397,1	466,2	536,5	607,7
14	152,1	221,7	293,8	367,8	443,6	520,8	599,4	679,0
15	168,6	245,7	325,7	407,8	491,9	577,4	664,6	752,8
16	185,6	270,6	358,6	449,1	541,7	635,9	732,0	829,2
17	203,1	296,2	392,6	491,7	593,1	696,3	801,5	907,9
18	221,2	322,6	427,6	535,5	646,1	758,5	837,1	989,0
19	239,8	349,7	463,3	580,6	700,5	822,4	946,7	1072,4
20	258,8	377,6	500,5	626,9	756,4	888,1	1022,3	1158,1
21	278,8	406,1	538,4	674,4	813,7	955,4	1099,8	1245,9
22	298,4	435,5	577,2	723,0	872,3	1024,3	1179,1	1335,7
23	318,9	465,3	616,9	772,7	932,4	1094,8	1260,3	1427,7
24	339,8	495,8	657,4	823,5	993,7	1166,8	1343,2	1521,7
25	361,1	527,0	698,8	875,4	1056,3	1240,4	1427,9	1617,6

Далее вычисляем разности $D_{l,m} = |R_l - R_m|$:

l	m		
	2	3	4
1	22,5	35,5	66,0
2		13,0	33,5
3			30,5

Так как $D_{1,4} = 66 > D_{l,m}(0{,}95) = 48{,}1$, гипотеза сдвига не отклоняется, между первой и четвертой выборками существует значимый сдвиг в параметрах положения.

4.2.1.2.3. Критерий Вилкоксона–Вилкокс

Критерий предложен в [454] и подобен критерию Неменьи (см. раздел 4.2.1.2.2). Пусть имеются k выборок равного объема n и x_{ij} — i-й элемент j-й выборки $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, k)$. Обозначим через R_{ij} ранг i-го наблюдения j-й выборки в упорядоченном по возрастанию ряду i-х элементов k выборок $(1 \leqslant R_{ij} \leqslant k)$ и через $R_j = \sum\limits_{i=1}^{n} R_{ij}$ сумму рангов j-й выборки.

Статистикой критерия является разность

$$D_{\nu,\varepsilon} = |R_\nu - R_\varepsilon|,\ (\nu \neq \varepsilon;\ \nu, \varepsilon = 1, 2, \ldots, k),$$

критические значения которой приведены в табл. 151.

При $D_{\nu,\varepsilon} \geqslant D_{\nu,\varepsilon}(\alpha)$ с доверительной вероятностью α гипотеза сдвига принимается.

Так же, как и критерий Неменьи, настоящий критерий позволяет выявить выборки, приводящие к отклонению нулевой гипотезы.

Задача 245. *Для данных задачи* 244 *проверить гипотезу сдвига критерием Викоксона–Вилкокс.*

Ранжируем первые элементы всех 4 выборок. Результаты сведем в таблицу:

i	j							
	1		2		3		4	
	x_{ij}	R_{ij}	x_{ij}	R_{ij}	x_{ij}	R_{ij}	x_{ij}	R_{ij}
1	1	1	3	2	4	3	7	4
2	2	1	4	2,5	4	2,5	8	4
3	3	1	5	2	6	3	9	4
4	4	1	6	2	7	3	10	4
5	5	1	7	2	8	3	11	4

Далее составляем разности $D_{\nu\varepsilon} = |R_\nu - R_\varepsilon|$:

ν	ε		
	2	3	4
1	5,5	9,5	15,0
2		4,0	9,5
3			5,5

Для $k = 4$, $n = 5$ и $\alpha = 0{,}95$ находим из табл. 151 $D_{\nu\varepsilon}(0{,}95) = 10{,}5$. Так как $D_{1,4} = 15 > D_{\nu\varepsilon}(0{,}95) = 10{,}5$, сдвиг между параметрами положения первой и четвертой выборок следует признать значимым (т. е. не случайным).

Таблица 151

Критические значения $D_{\nu,\varepsilon}(\alpha)$ критерия Вилкоксона–Вилкокс [9]

n	k							
	3	4	5	6	7	8	9	10
Доверительная вероятность $\alpha = 0{,}90$								
3	5,0	7,2	9,5	11,9	14,2	16,7	19,1	21,7
4	5,8	8,4	11,0	13,7	16,5	19,3	22,1	25,0
5	6,5	9,4	12,3	15,3	18,4	21,5	24,7	28,0
6	7,1	10,2	13,5	16,8	20,2	23,6	27,1	30,6
7	7,7	11,1	14,5	18,1	21,8	25,5	29,3	33,1
8	8,2	11,8	15,6	19,4	23,3	27,2	31,3	35,4
9	8,7	12,5	16,5	20,5	24,7	28,9	33,2	37,3
10	9,2	13,2	17,4	21,7	26,0	30,4	35,0	41,5
11	9,6	13,9	18,2	22,7	27,3	31,9	36,7	43,3
12	10,1	14,5	19,0	23,7	28,5	33,4	38,3	44,3
13	10,5	15,1	19,8	24,7	29,7	34,7	39,9	45,1
14	10,9	15,7	20,6	25,6	30,8	36,0	41,4	46,8
15	11,2	16,2	21,3	26,5	31,9	37,3	42,8	48,4
16	11,6	16,7	22,0	27,4	32,9	38,5	44,2	50,0
17	12,0	17,2	22,7	28,2	33,9	39,7	45,6	51,5
18	12,3	17,7	23,3	29,1	34,9	40,9	46,9	53,0
19	12,6	18,2	24,0	29,9	35,9	42,0	48,2	54,5
20	13,0	18,7	24,6	30,6	36,9	43,1	49,4	55,9
21	13,3	19,2	25,2	31,4	37,7	44,1	50,7	57,3
22	13,6	19,6	25,9	32,1	38,6	45,2	51,9	58,6
23	13,9	20,1	26,4	32,8	39,5	46,2	53,0	60,0
24	14,2	20,5	26,9	33,6	40,5	47,2	54,2	61,2
25	14,5	20,9	27,5	34,2	41,1	48,1	55,3	62,5
Доверительная вероятность $\alpha = 0{,}95$								
3	5,7	8,1	10,6	13,1	15,6	18,2	20,8	23,5
4	6,6	9,4	12,2	15,1	18,0	21,0	24,0	27,1
5	7,4	10,5	13,6	16,9	20,1	23,5	26,9	30,3
6	8,1	11,5	14,9	18,5	22,1	25,7	29,4	33,2
7	8,8	12,4	16,1	19,9	23,9	27,8	31,8	35,8
8	9,4	13,3	17,3	21,3	25,5	29,7	34,0	38,3
9	9,9	14,1	18,3	22,6	27,0	31,5	36,0	40,6
10	10,5	14,8	19,3	23,8	28,5	33,2	38,0	42,8
11	11,0	15,6	20,2	25,0	29,9	34,8	39,8	44,9
12	11,5	16,2	21,1	26,1	31,2	36,4	41,6	46,9
13	11,9	16,9	22,0	27,2	32,5	37,9	43,3	48,8
14	12,4	17,5	22,8	28,2	33,7	39,3	45,0	50,7
15	12,8	18,2	23,6	29,2	34,9	40,7	46,5	52,5
16	13,3	18,8	24,4	30,2	36,0	42,0	48,1	54,2
17	13,7	19,3	25,2	31,1	37,1	43,3	49,5	55,9
18	14,1	19,9	25,9	32,0	38,2	44,5	51,0	57,5
19	14,4	20,4	26,6	32,9	39,3	45,8	52,4	59,0
20	14,8	21,0	27,3	33,7	40,3	47,0	53,7	60,6
21	15,2	21,5	28,0	34,6	41,3	48,1	55,1	62,1
22	15,5	22,0	28,6	35,4	42,3	49,2	56,4	63,5
23	15,9	22,5	29,3	36,2	43,2	50,3	57,6	65,0
24	16,2	23,0	29,9	36,9	44,1	51,4	58,9	66,4
25	16,6	23,5	30,5	37,7	45,0	52,5	60,1	67,7

4.2.1.2.4. „Быстрый“ критерий Кенуя

Среди k выборок равного объема $n \geqslant 20$ находятся наибольшее среди наименьших значений $x_{\max}^{\min}$ и наименьшее среди наибольших значений $x_{\min}^{\max}$ в выборках. Подсчитываются количества n_1 наблюдений, для которых $x_i < x_{\max}^{\min}$, и n_2, для которых $x_i < x_{\min}^{\max}$.

Статистикой критерия Кенуя является сумма $n = n_1 + n_2$, критические значения которой $n(\alpha)$ приведены в табл. 152.

Таблица 152

Критические значения $n(\alpha)$ критерия Кенуя
(α — уровень значимости) [121]

k	2	3	4	5	6	8	10
$\alpha = 0{,}05$	9	17	27	37	47	70	93
$\alpha = 0{,}01$	12	22	33	45	57	83	110

Задача 246. *Имеются три выборки случайных величин объема $n = 20$ каждая*

x_{1j}: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39;
x_{2j}: 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48;
x_{3j}: 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40.

Необходимо проверить гипотезу сдвига критерием Кенуя на уровне значимости $\alpha = 0{,}05$.

Имеем $x_{\max}^{\min} = 21$ и $x_{\min}^{max} = 39$. Тогда $n_1 = 16$, $n_2 = 6$ и $n = 16 + 6 = 22$.
Из табл. 152 для $k = 3$ и $\alpha = 0{,}05$ находим $n(\alpha) = 17$.
Так как $n = 22 > n(\alpha) = 17$, гипотеза сдвига не отклоняется.

4.2.1.2.5. Критерий Фишера–Терри–Йэйтса–Гёфдинга

Вариант критерия Фишера–Терри–Йэйтса–Гёфдинга (см. раздел 4.2.1.1.2.3) для $k > 2$ выборок основан на статистике [365]

$$Q = \left\{\sum_{i=1}^{N} a_N^2(i)\right\}^{-1} (N-1) \sum_{j=1}^{k} \frac{1}{n_j} \left[\sum_{i=1}^{N} a_N(R_j)\right]^2,$$

где R_{ij} — ранг элементов j-й выборки в общем ряду ($N = \sum_{j=1}^{k} n_j$).

При $n_j > 10$ распределение статистики Q может быть аппроксимировано χ^2-распределением с $f = k - 1$ степенями свободы. Поэтому нулевая гипотеза отсутствия сдвига отклоняется с достоверностью α, если $Q > \chi_\alpha^2(k-1)$.

Задача 247. *Имеются три выборки случайных величин:*

x_{1j}: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10;
x_{2j}: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25;
x_{3j}: 12, 14, 16, 18, 20, 22, 24, 26, 28, 30.

Необходимо проверить гипотезу сдвига критерием Фишера–Терри–Йэйтса–Гёфдинга при доверительной вероятности $\alpha = 0{,}95$.

Имеем $n_1 = 10$, $n_2 = 11$, $n_3 = 10$ и $N = 31$.

Строим общий, ранжированный по возрастанию, ряд из элементов трех заданных выборок ($i = 1, 2, 3$ — номер первичной выборки, x_j — элементы объединенной выборки, $j = 1, 2, \ldots, 31$; R_{ij} — ранг j-го элемента в объединенной выборке):

i	x_{ij}	R_{ij}	i	x_{ij}	R_{ij}	i	x_{ij}	R_{ij}	i	x_{ij}	R_{ij}
1	1	1	2	7	8,5	3	14	17	3	22	25
1	2	2	1	8	10	2	15	18	2	23	26
1	3	3	1	9	11,5	3	16	19	3	24	27
1	4	4	2	9	11,5	2	17	20	2	25	28
1	5	5,5	1	10	13	3	18	21	3	26	29
2	5	5,5	2	11	14	2	19	22	3	28	30
1	6	7	3	12	15	3	20	23	3	30	31
1	7	8,5	2	13	16	2	21	24			

Далее вычисляем значения (см. раздел 4.2.1.1.2.3)

$$a_{31}(k) = 4{,}91 \cdot \left[\left(\frac{k - \frac{3}{8}}{31 + \frac{1}{4}} \right)^{0,14} - \left(1 - \frac{k - \frac{3}{8}}{31 + \frac{1}{4}} \right)^{0,14} \right],$$

включая дробные значения k (дробные ранги).

Результаты сведем в таблицу:

k	$a_{31}(k)$	k	$a_{31}(k)$	k	$a_{31}(k)$	k	$a_{31}(k)$
1	−2,0567	8,5	−0,6412	16	0	25	0,7974
2	−1,6276	9	−0,5927	17	0,0799	26	0,9134
3	−1,3788	10	−0,4996	18	0,1604	27	1,0434
4	−1,1943	11	−0,4108	19	0,2419	28	1,1943
5	−1,0434	11,5	−0,3676	20	0,3252	29	1,3788
5,5	−0,97634	12	−0,3252	21	0,4108	30	1,6276
6	−0,9134	13	−0,2419	22	0,4996	31	2,0567
7	−0,7974	14	−0,1604	23	0,5927		
8	−0,6913	15	−0,0799	24	0,6913		

Для первой выборки ($i = 1$) имеем

$$R_{1j}\colon 1,\ 2,\ 3,\ 4,\ 5{,}5,\ 7,\ 8{,}5,\ 10,\ 11{,}5,\ 13 \quad \text{и} \quad \sum_{j=1}^{31} a_{31}(R_{1j}) = -9{,}7814.$$

Для второй ($i = 2$) выборки

$$R_{2j}\colon 5{,}5,\ 8{,}5,\ 11{,}5,\ 14,\ 16,\ 18,\ 20,\ 22,\ 24,\ 26,\ 28 \quad \text{и} \quad \sum_{j=1}^{31} a_{31}(R_{2j}) = 1{,}6387.$$

Для третьей выборки ($i = 3$)

$$R_{3j}\colon 15,\ 17,\ 19,\ 21,\ 23,\ 25,\ 27,\ 29,\ 30,\ 31 \quad \text{и} \quad \sum_{j=1}^{31} a_{31}(R_{3j}) = 8{,}1293.$$

Находим $\sum_{i=1}^{31} a_{31}^2(i) = 28{,}4186$ и вычисляем статистику критерия

$$Q = \frac{3-1}{28{,}4186} \cdot \sum_{j=1}^{3} \frac{1}{n_j} \cdot \left[\sum_i a_n(R_{ij}) \right]^2 = 17{,}367.$$

Находим из табл. 55: $\chi^2_{0,95}(2) = 5{,}99$.

Так как $Q = 17{,}367 > \chi^2_{0,95}(2) = 5{,}99$, гипотеза сдвига не отклоняется.

4.2.1.2.6. Критерий Ван дер Вардена

Статистика Ван дер Вардена для $k > 2$ выборок имеет вид

$$Q^* = \left\{\sum_{i=1}^{N} u^2_{\frac{i}{N+1}}\right\}^{-1} (N-1)\sum_{j=1}^{k}\frac{1}{n_j}\left[\sum_{i=1}^{N} a_N(R_{ij})\right]^2.$$

При справедливости нулевой гипотезы статистика Q^* распределена как χ^2 с $f = k-1$ степенями свободы. Поэтому нулевая гипотеза отсутствия сдвига отклоняется, если $Q^* > \chi^2_\alpha(k-1)$, где α — доверительная вероятность.

Задача 248. *Проверить гипотезу сдвига в условиях задачи 247 критерием Ван дер Вардена.*

Вычисляем $u_{\frac{i}{N+1}} = 4{,}91\cdot\left[\left(\frac{i}{32}\right)^{0,14} - \left(1-\frac{i}{32}\right)^{0,14}\right]$.

Результаты сведем в таблицу (включая дробные значения рангов i):

i	$u_{\frac{i}{32}}$	i	$u_{\frac{i}{32}}$	i	$u_{\frac{i}{32}}$	i	$u_{\frac{i}{32}}$
1	−1,8658	8,5	−0,6723	16	0	24	0,6723
2	−1,5354	9	−0,5771	17	0,078	25	0,7743
3	−1,3178	10	−0,4869	18	0,1566	26	0,8851
4	−1,1492	11	−0,4006	19	0,2362	27	1,0083
5	−1,0083	11,5	−0,3586	20	0,3173	28	1,1492
5,5	−0,9448	12	−0,3173	21	0,4006	29	1,3178
6	−0,8851	13	−0,2362	22	0,4869	30	1,5354
7	−0,7743	14	−0,1566	23	0,5771	31	1,8658
8	−0,6723	15	−0,078				

Легко видеть, что соотношение $u_{\frac{i}{N+1}} = -u_{\frac{N+1-i}{N+1}}$ позволяет ограничиться расчетами для $i < \dfrac{N+1}{2}$. Из табл. 147 находим

$$K = \frac{1}{31}\cdot\sum_{i=1}^{31} u^2_{\frac{i}{32}} = 0{,}817;\qquad \sum_{i=1}^{31} u^2_{\frac{i}{N+1}} = 0{,}817\cdot 31 = 25{,}327.$$

Находим последовательно для $i = 1, 2, 3$:

$$\sum_{j=1}^{31} u_{\frac{R_{1j}}{N+1}} = -9{,}293;\qquad \sum_{j=1}^{31} u_{\frac{R_{2j}}{N+1}} = 1{,}5834;\qquad \sum_{j=1}^{31} u_{\frac{R_{3j}}{N+1}} = 6{,}1801.$$

Окончательно получаем

$$Q = \frac{30}{25{,}327}\cdot\sum_{j=1}^{3}\frac{1}{n_j}\cdot\left(u_{\frac{R_{ij}}{N+1}}\right)^2 = 18{,}3848.$$

Так как $Q^* = 18{,}38 > \chi^2_{0,95}(2) = 5{,}99$, гипотеза сдвига принимается, как и ранее.

4.2.1.2.7. Медианный критерий

Для множественного аналога двухвыборочного медианного критерия (см. раздел 4.2.1.1.2.5) используется статистика

$$\chi^2 = 4\sum_{i=1}^{k}\frac{1}{n_i}m_i^2 - N,$$

имеющая при $n_i \to \infty$ (> 10) распределение хи-квадрат с $f = k - 1$ степенями свободы. Здесь m_i — число наблюдений i-й выборки, превосходящих медиану объединенной выборки ($N = \sum\limits_{i=1}^{k} n_i$).

Задача 249. *Проверить гипотезу сдвига в условиях задачи* 247 *медианным критерием.*

Медианой объединенной выборки является $\tilde{x} = 13$.
Имеем $m_1 = 0$, $m_2 = 6$, $m_3 = 9$, $N = 31$. Тогда

$$\chi^2 = 4 \cdot \sum_{i=1}^{3} \frac{1}{n_i} \cdot m_i^2 - 31 = 4 \cdot \left(\frac{0^2}{10} + \frac{6^2}{10} + \frac{9^2}{10}\right) - 31 = 14{,}492.$$

Так как $\chi^2 = 14{,}492 > \chi^2_{0,95}(2) = 5{,}99$, гипотеза сдвига не отклоняется.

4.2.1.2.8. Критерий Хеттманспергера

Критерий рассмотрен в [455]. Используется для проверки равенства параметров положения $H_0\colon\ \mu_1 = \ldots = \mu_k$ против альтернатив упорядоченности $H_1\colon\ \mu_1 \leqslant \ldots \leqslant \mu_k$, где хотя бы одно из неравенств — строгое. Статистика критерия

$$L = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \left(i - \frac{k+1}{2}\right)\left(\bar{R}_i - \frac{N+1}{2}\right).$$

Нулевая гипотеза отклоняется с достоверностью α, если

$$L \geqslant u_\alpha \sqrt{\mathbf{D}(L)},$$

где u_α — α-квантиль стандартного нормального распределения,

$$\mathbf{D}(L) = \frac{N+1}{2} \sum_{i=1}^{k} \frac{1}{n_i} \left(i - \frac{k+1}{2}\right)^2,$$

$$\text{при } n_i = n \quad (i = 1, 2, \ldots, k) \quad \mathbf{D}(L) = \frac{\left(k^2 - 1\right)(n - k + 1)}{144n}.$$

Как и в критерии Крускала–Уоллиса, приняты обозначения $\bar{R}_i = \dfrac{R_i}{n}$, $R_i = \sum\limits_{j=1}^{N} R_{ij}$ и R_{ij} — ранг j-го элемента i-й выборки в общем упорядоченном ряду.

Задача 250. *Проверить гипотезу сдвига против порядковой альтернативы критерием Хеттманспергера в условиях задачи* 243.

Имеем (см. задачу 243)

$$\bar{R}_1 = 5{,}333; \quad \bar{R}_2 = 9{,}1; \quad \bar{R}_3 = 16{,}166; \quad \bar{R}_4 = 13{,}75; \quad \bar{R}_5 = 23{,}64.$$

Тогда

$$L = \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{k} \left(i - \frac{k+1}{2}\right) \cdot \left(\bar{R}_i - \frac{N+1}{2}\right) = \frac{1}{\sqrt{27}} \cdot \sum_{i=1}^{5} (i-3) \cdot \left(\bar{R}_i - 14\right) =$$
$$= 0{,}1924 \cdot [(1-3)\cdot(5{,}333-14) + (2-3)\cdot(9{,}1-14) + \ldots + (5-3)\cdot(23{,}64-14)] = 7{,}94;$$

$$\mathbf{D}(L) = \frac{28}{12} \cdot \sum_{i=1}^{5} \frac{1}{n_i} \cdot (i-3)^2 = 3{,}7439 \quad \left(\sqrt{\mathbf{D}(L)} = 1{,}935\right).$$

Окончательно имеем для $\alpha = 0{,}95$: $u_{0,95} = 1{,}645$ и $u_{0,95} \cdot \sqrt{\mathbf{D}\,(L)} = 3{,}183$.
Так как $L = 7{,}94 > 3{,}183$, гипотеза сдвига не отклоняется.

4.2.1.2.9. Критерий Терпстры–Джонкхира

Так же как и критерий Хеттманспергера (см. раздел 4.2.1.2.8), используется для проверки гипотезы сдвига против альтернатив упорядоченности. Критерий предложен в [456, 457] и основан на попарных статистиках Вилкоксона–Манна–Уитни (см. раздел 4.2.1.1.2.2). Статистика критерия имеет вид

$$S = \sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij},$$

где a_{ij} — количество наблюдений из первых $(i-1)$ выборок $(i > 1)$, меньших, чем x_{ij} — j-е наблюдение в i-й выборке.

Гипотеза отсутствия сдвига отклоняется, если $S > S_\alpha$, где S_α — критические значения, приведенные в табл. 153 и 154.

При $n_i \geqslant 10$ применима аппроксимация

$$S_\alpha = \mathbf{M}(S) + u_\alpha \mathbf{D}(S),$$

где

$$\mathbf{D}(S) = \frac{1}{72}\left\{\left(\sum_{i=1}^{k} n_i\right)^2 \left(2\sum_{i=1}^{k} n_i + 3\right) - \sum_{i=1}^{k} n_i^2\,(2n_i + 3)\right\};$$

u_α — α-квантиль стандартного нормального распределения.

Легко убедиться, что статистики критериев L — Хеттманспергера и S — Терпстры–Джонкхира эквивалентны.

Задача 251. *Проверить гипотезу сдвига против порядковой альтернативы критерием Терпстры–Джонкхира для данных задачи* 243.

Поясним порядок вычисления критерия S. Для $i = 2$ (второй выборки) находим количество наблюдений первой выборки, меньших, чем x_{2j}.

Для $j = 1$ имеем $x_{21} = 3$ и в первой выборке два значения ($x_{11} = 1$ и $x_{12} = 2$) меньше, чем $x_{21} = 3$, т. е. $a_{21} = 2$.

Далее по аналогии для $j = 2$ имеем $a_{22} = 3$ (предлагается убедиться в этом самому читателю). Для $j = 3$ имеем $a_{23} = 4$ и далее $a_{24} = 5$ и $a_{25} = 6$.

При $i = 3$ находим количества значений в первых двух выборках, меньших, чем x_{3j} (при различных j): $a_{31} = 10$, $a_{32} = 11$, $a_{33} = 11$.

Для $i = 4$ имеем $a_{41} = 0$, $a_{42} = 6$, $a_{43} = 10$, $a_{44} = 12$, $a_{45} = 14$, $a_{46} = 14$.

Для $i = 5$ имеем $a_{51} = 18$, $a_{52} = 19$, $a_{53} = 20$, $a_{54} = 20$, $a_{55} = 20$, $a_{56} = 20$, $a_{57} = 20$.

Окончательно находим

$$S = \sum_{i=1}^{5} \sum_{j=1}^{n_i} a_{ij} = 2 + 3 + 4 + 5 + 6 + 10 + 11 + 11 + \ldots + 20 + 20 = 245.$$

Далее вычисляем

$$\mathbf{M}(S) = \frac{(6+5+3+6+7)^2 - (6^2+5^2+3^2+6^2+7^2)}{4} = 143{,}5;$$

$$\mathbf{D}(S) = \frac{1}{72} \cdot [729 \cdot (2 \cdot 27 + 3) - 36 \cdot (2 \cdot 6 + 3) - 25 \cdot (2 \cdot 3 + 3) - \ldots$$

$$\ldots - 49 \cdot (2 \cdot 7 + 3)] = 344{,}916; \quad \sqrt{\mathbf{D}(S)} = 23{,}343.$$

Находим (имея в виду, что $u_{0,95} = 1{,}645$) $S_{0,95} = 143{,}5 + 1{,}645 \cdot 23{,}343 = 181{,}9$.

Так как $S = 245 > S_{0,95} = 181{,}9$, имеющиеся результаты наблюдений не противоречат гипотезе сдвига.

Таблица 153

Критические значения S_α критерия Терпстры–Джонкхира для $k = 3$
(α — доверительная вероятность) [458]

n_1	n_2	n_3	α		n_1	n_2	n_3	α		n_1	n_2	n_3	α	
			0,90	0,95				0,90	0,95				0,90	0,95
2	2	2	9	20	3	3	3	19	21	4	5	7	55	58
2	2	3	12	21	3	3	4	23	25	4	5	8	60	64
2	2	4	15	16	3	3	5	27	29	4	6	6	55	59
2	2	5	18	19	3	3	6	31	33	4	6	7	62	66
2	2	6	20	22	3	3	7	35	38	4	6	8	68	72
2	2	7	23	25	3	3	8	39	42	4	7	7	68	73
2	2	8	26	28	3	4	4	28	30	4	7	8	75	80
2	3	3	15	17	3	4	5	32	35	4	8	8	82	87
2	3	4	19	20	3	4	6	37	39	5	5	5	50	53
2	3	5	22	24	3	4	7	41	44	5	5	6	56	60
2	3	6	25	27	3	4	8	46	49	5	5	7	62	66
2	3	7	29	31	3	5	5	37	40	5	5	8	68	73
2	3	8	32	35	3	5	6	42	45	5	6	6	63	67
2	4	4	23	25	3	5	7	48	51	5	6	7	69	74
2	4	5	27	29	3	5	8	53	56	5	6	8	76	81
2	4	6	31	33	3	6	6	48	51	5	7	7	77	81
2	4	7	35	37	3	6	7	54	57	5	7	8	84	89
2	4	8	38	41	3	6	8	57	63	5	8	8	92	97
2	5	5	31	34	3	7	7	60	64	6	6	6	70	74
2	5	6	36	38	3	7	8	66	70	6	6	7	77	82
2	5	7	40	43	3	8	8	73	778	6	6	8	85	90
2	5	8	45	48	4	4	4	33	35	6	7	7	85	90
2	6	6	41	44	4	4	5	38	41	6	7	8	93	99
2	6	7	46	49	4	4	6	43	46	6	8	8	101	107
2	6	8	51	54	4	4	7	48	51	7	7	7	94	99
2	7	7	52	55	4	4	8	53	57	7	7	8	102	108
2	7	8	57	61	4	5	5	44	47	7	8	8	111	117
2	8	8	63	68	4	5	6	49	53	8	8	8	120	127

Таблица 154

Критические значения S_α критерия Терпстры–Джонкхира для выборок равного объема
(α — доверительная вероятность) [458]

n	k	α		n	k	α	
		0,90	0,95			0,90	0,95
2	4	17	18	4	6	146	153
	5	27	29	5	4	94	99
	6	39	42		5	152	159
3	4	36	39		6	223	233
	5	58	61	6	4	133	140
	6	84	89		5	215	225
4	4	62	66		6	316	329
	5	99	105				

4.2.1.2.10. Критерий Мостеллера

Для k выборок равного объема n критерий позволяет ответить на вопрос: не является ли одна из k выборок сдвинутой по отношению к $(k-1)$ остальным?

Вероятность того, что в одной из k выборок (равного объема n) r или более членов больше, чем в $(k-1)$ оставшихся выборках, равна [455] $\mathbf{P}(r) = kC_{kn-r}^{n-r}(C_{kn}^{n})^{-1}$ или при $n \to \infty$ и фиксированных k и r:

$$\mathbf{P}(r) = \frac{1}{k^{r-1}}\left[1 - \frac{r(2r-1)(k-1)}{2kn}\right].$$

Если $\mathbf{P}(r) < 1-\alpha$, где α — доверительная вероятность, то гипотеза сдвига отклоняется.

Задача 252. *Для данных задачи 244 проверить гипотезу сдвига критерием Мостеллера.*

Проверяем четвертую выборку. В ней три члена (9, 10 и 11) превышают все значения остальных 3 выборок. Имеем $k=4$, $n=5$ и $r=3$. Вычисляем

$$\mathbf{P}(3) = 4\cdot\frac{C_{4\cdot5-3}^{5-3}}{C_{4\cdot5}^{5}} = 4\cdot\frac{C_{17}^{2}}{C_{20}^{5}} = 4\cdot\frac{136}{15504} = 0,035.$$

Так как $\mathbf{P}(3) = 0,035 < 1-\alpha = 0{,}05$, сдвиг четвертой выборки по отношению ко всем остальным следует признать значимым.

4.2.1.2.11. Критерий Левиса

Критерий аналогичен критерию Мостеллера [459] и использует соотношение между размахами сравниваемых выборок.

Пусть для i-й выборки имеем $\min\limits_{1\leqslant j\leqslant n} x_{ij} = U_i$, $\max\limits_{1\leqslant j\leqslant n} x_{ij} = V_i$ и $R_i = V_i - U_i$ — размах i-й выборки.

Обозначим $\max U_i = U$ и $\min V_i = V$. Параметры положения сравниваемых выборок не равны между собой (гипотеза H_0 отклоняется), если по крайней мере в двух выборках размахи не пересекаются, т. е. если $U > V$.

Гипотеза сдвига принимается с достоверностью α, если [459]

$$\mathbf{P}(U>V) < 1-\alpha,$$

где

$$\mathbf{P}(U>V) = \sum_{i=2}^{k}\sum_{j=1}^{i-1}(-1)^i\frac{C_k^i C_i^j}{C_{in}^{jn}}.$$

При $n\to\infty$ $(n>10)$

$$\mathbf{P}(U>V) = k(k-1)\frac{(n!)^2}{(2n)!}.$$

Задача 253. *Для данных задачи 244 проверить гипотезу сдвига критерием Левиса.*

Имеем $U_1=1$, $U_2=3$, $U_3=4$, $U_4=7$; $V_1=5$, $V_2=7$, $V_3=8$, $V_4=11$; $U=\max U_i=7$; $V=\min V_i=5$.

Находим

$$\mathbf{P}(U>V) = \sum_{i=2}^{4}\sum_{j=1}^{i-1}(-1)^i\cdot\frac{C_4^i C_i^j}{C_{5i}^{5j}} =$$
$$= \sum_{j=1}^{1}(-1)^2\cdot\frac{C_4^2C_4^j}{C_{10}^{5j}} + \sum_{j=1}^{2}(-1)^3\cdot\frac{C_4^3C_3^j}{C_{15}^{5j}} + \sum_{j=1}^{3}(-1)^4\cdot\frac{C_4^4C_4^j}{C_{20}^{5j}} = 0{,}04.$$

Так как $\mathbf{P}(U>V) = 0{,}04 < 1-0{,}95 = 0{,}05$, гипотеза сдвига не отклоняется.

4.2.1.2.12. L-критерий, основанный на U-статистиках

Пусть имеются k выборок объема n_i $(i = 1, \ldots, k)$ каждая. Из каждой выборки берем по одному наблюдению. Проделав это всеми возможными способами, в результате получаем $\prod\limits_{i=1}^{k} n_i$ выборок одинакового объема k.

Обозначим через ν_{ij} (ν_{ik}) количество выборок, в которых наблюдение, взятое из i-й выборки, было наименьшим (наибольшим). Определим величину $u_{ij} = \dfrac{\nu_{ij}}{\prod\limits_i n_i}$ $\left(u_{ik} = \dfrac{\nu_{ik}}{\prod\limits_i n_i}\right)$.

В [461, 462] предложен критерий сдвига, основанный на статистике

$$L = \frac{N(2k-1)(k-1)C_{2k-2}^{k-1}}{2k^2(C_{2k-2}^{k-1}-1)} \left\{ \sum_{i=1}^{k} \frac{n_i}{N} l_i^2 - \left[\sum_{i=1}^{k} \frac{n_i}{N} l_i \right]^2 \right\},$$

где $l_i = -u_{ij} + u_{ik}$.

При $L > L_\alpha$ гипотеза сдвига принимается с достоверностью α (здесь L_α — критическое значение, приведенное в табл. 155).

Таблица 155

Критические значения L_α для $k = 3$ и доверительной вероятности $\alpha = 0{,}95$ [460]

n_1	n_2	n_3	$L_{0,95}$	n_1	n_2	n_3	$L_{0,95}$
2	2	2	5,33	2	4	4	5,55
2	2	3	5,39	3	3	3	6,22
2	2	4	6,00	3	3	4	6,16
2	3	3	5,85	3	4	4	6,22
2	3	4	5,78	4	4	4	6,12

При $k > 3$ справедливо приближение

$$L_\alpha \approx \chi^2_\alpha(f - k - 1),$$

где $\chi^2_\alpha(f)$ — α-квантиль распределения хи-квадрат с f степенями свободы.

Задача 254. *Даны три выборки случайных величин е*

$$x_{1j}\colon 1,\ 2,\ 3; \quad x_{2j}\colon 4,\ 5,\ 6; \quad x_{3j}\colon 7,\ 8,\ 9.$$

Проверить гипотезу сдвига L-критерием при доверительной вероятности $\alpha = 0{,}95$ $(k = 3,\ n_i = 3)$.

В нашем случае следует рассмотреть $\prod\limits_i n_i = 3 \cdot 3 \cdot 3 = 27$ выборок равного объема $n = 3$.

Имеем выборки:

1,4,7	1,5,7	1,6,7	2,4,7	2,5,7	2,6,7	3,4,7	3,5,7	2,6,7
1,4,8	1,5,8	1,6,8	2,4,8	2,5,8	2,6,8	3,4,8	3,5,8	2,6,8
1,4,9	1,5,9	1,6,9	2,4,9	2,5,9	2,6,9	3,4,9	3,5,9	3,6,9

Находим при $i = 1$ (первая выборка) $v_{11} = 27$, $v_{13} = 0$ и по аналогии $v_{21} = 0$, $v_{23} = 0$, $v_{31} = 0$, $v_{33} = 27$.

Следовательно, имеем

$$u_{11} = \frac{\nu_{11}}{27} = 1; \quad u_{13} = 0; \quad u_{21} = 0; \quad u_{23}0; \quad u_{31} = 1;$$
$$l_1 = -1 + 0 = -1; \quad l_2 = -0 + 0 = 0; \quad l_3 = -0 + 1 = 1;$$
$$L = \frac{9 \cdot 5 \cdot 4 \cdot C_4^2}{2 \cdot 9 \cdot (C_4^2 - 1)} \cdot \left\{ \sum_{i=1}^{3} \frac{3}{9} \cdot l_i^2 - \left(\sum_{i=1}^{3} \frac{3}{9} \cdot l_i \right)^2 \right\} = 8.$$

Из табл. 155 для $n_1 = n_2 = n_3 = 3$ находим $L_{0,95} = 6{,}22$; так как $L = 8 > L_{0,95} = 6{,}22$, гипотеза сдвига не отклоняется.

4.2.1.2.13. Критерий Краузе

Этот критерий является обобщением критерия Вилкоксона–Манна–Уитни (см. раздел 4.2.1.1.2.2) на многовыборочный случай. К этому классу относятся критерии Крускала–Уоллиса (см. раздел 4.2.1.2.1) и Терпстры–Джонкхира (см. раздел 4.2.1.2.9).

Рассмотрим формы обобщения критерия Вилкоксона–Манна–Уитни, порождающие различные многовыборочные критерии.

Пусть W_{ij} — статистика Манна–Уитни, возникающая при сравнении выборок с номерами i и j, т. е. (см. раздел 4.2.1.1.2.2) W_{ij} есть точное число пар значений $x_{i\nu}$ и $x_{j\varepsilon}$, для которых $x_{i\nu} < x_{j\varepsilon}$ по всем $\nu = 1, 2, \ldots, n_i$ и $\varepsilon = 1, 2, \ldots, n_j$.

Тогда $W_i = \sum\limits_{i \neq j} W_{ij}$ будет статистикой Манна–Уитни, сравнивающей i-ю выборку со всеми остальными. Можно показать [463], что

$$W_i = n_i \left[N - \frac{1}{2}(n_i - 1) - \bar{R}_i \right],$$

где $N = \sum\limits_{i=1}^{k} n_i$ и $\bar{R}_i$ — среднее из рангов членов i-й выборки в общем ранжированном ряду $N = \sum n_i$ значений.

При справедливости нулевой гипотезы (отсутствие сдвига в параметрах положения k выборок) математическое ожидание $\mathbf{M}(W_i)$ и дисперсия $\mathbf{D}(W_i)$ равны

$$\mathbf{M}(W_i) = \frac{1}{2} n_i (N - n_i); \quad \mathbf{D}(W_i) = \frac{1}{12} n_i (N - n_i)(N + 1).$$

В принятых обозначениях рассмотренные ранее статистики H — Крускала–Уоллиса (см. раздел 4.2.1.2.1) и S — Терпстры–Джонкхира (см. раздел 4.2.1.2.9) могут быть записаны как

$$H = \frac{1}{N} \sum_{i=1}^{k} (N - n_i) \frac{[W_i - \mathbf{M}(W_i)]^2}{\mathbf{D}(W_i)}; \quad S = 2 \sum_{1 \leqslant i < j \leqslant k} W_{ij} - \sum_{1 \leqslant i < j \leqslant k} n_i n_j.$$

В этом разделе мы рассматриваем статистику критерия, предложенного Краузе [464, 465]:

$$V = \frac{12}{N+1} \sum_{1 \leqslant i < j \leqslant k} \frac{\left(W_{ij} - \frac{n_i n_j}{2} \right)^2}{n_i n_j}.$$

Этот критерий эквивалентен критерию Крускала–Уоллиса, но более эффективен против большого количества альтернатив.

При справедливости нулевой гипотезы статистика V распределена как χ^2 с $f = k - 1$ степенями свободы. Поэтому гипотеза сдвига отклоняется с достоверностью α, если $V < \chi^2_\alpha(k-1)$.

Задача 255. *Проверить гипотезу сдвига для данных задачи 243 непараметрическим критерием Краузе.*

Для выборок $i = 1$ и $j = 2$. Имеем $W_{12} = 5+5+4+3+2+1 = 20$.

Поясним способ подсчета W_{ij}. Для первого элемента выборки с номером $i = 1$, равного 1 во второй выборке имеются 5 элементов (3, 4, 5, 6, 7), превосходящих его. Для 2-го элемента выборки, равного 2, во второй выборке находятся также 5 элементов, превосходящих его по величине. Для 3-го элемента выборки, равного 3, их во второй выборке уже 4, и т. д.

По аналогии имеем

$W_{13} = 3+3+3+3+3+3+3 = 18$; $W_{14} = 5+5+5+5+4+4 = 28$;

$W_{15} = 7+7+7+7+7+7 = 42$; $W_{23} = 3+3+3+3+2 = 14$;

$W_{24} = 5+5+4+4+4+3 = 21$; $W_{25} = 7+7+7+7+7 = 35$;

$W_{34} = 3+2+2 = 7$; $W_{35} = 7+7+7 = 21$; $W_{45} = 7+7+7+7+6+5 = 39$.

Вычислим теперь

$$V = \frac{12}{27+1}\cdot\left\{\frac{\left(20-\dfrac{6\cdot 5}{2}\right)^2}{6\cdot 5}+\frac{\left(18-\dfrac{6\cdot 3}{2}\right)^2}{6\cdot 3}+\ldots+\frac{\left(39-\dfrac{6\cdot 7}{2}\right)^2}{6\cdot 7}\right\} = 19{,}1.$$

Из табл. 55 находим $\chi^2_{0,95}(a = 5-1 = 4) = 9.49$.

Так как $V = 19.1 > \chi^2_{0,95}(4) = 9.49$, гипотеза сдвига не отклоняется.

4.2.1.2.14. Критерий Пейджа

Используется для выборок равного объема в целях проверки гипотезы отсутствия сдвига $H_0\colon \mu_1 = \ldots = \mu_k$ против порядковых альтернатив $H_1\colon \mu_1 \leqslant \mu_2 \leqslant \ldots \leqslant \mu_k$, так же как критерии Хеттманспергера (см. раздел 4.2.1.2.8) и Терпстры–Джонкхира (см. раздел 4.2.1.2.9). Статистика критерия строится следующим образом. Составляется таблица:

Номер элемента в выборке	Номер выборки			
	1	2	...	k
1	x_{11}	x_{21}	...	x_{k1}
2	x_{12}	x_{22}	...	x_{k2}
$\vdots$	$\vdots$		...	$\vdots$
n	x_{1n}	x_{2n}	...	x_{kn}

Далее ранжируем по возрастанию элементы каждой строки от 1 до k (получаем совокупность рангов r_{ij}). Обозначим через $R_i = \sum_{j=1}^{n} r_{ij}$ сумму рангов по столбцам. Тогда статистика критерия Пейджа записывается в форме [466]

$$L = \sum_{i=1}^{k} iR_i.$$

Нулевая гипотеза отклоняется с достоверностью α, если $L \geqslant L_\alpha(k,n)$, где $L_\alpha(k,n)$ — критические значения, приведенные в табл. 156.

Таблица 156

Критические значения $L_\alpha(k, n)$ критерия Пейджа для доверительной вероятности $\alpha = 0{,}95$ [18]

n	k					
	3	4	5	6	7	8
2	28	58	103	166	252	362
3	41	84	150	244	370	532
4	54	111	197	321	487	701
5	66	137	244	397	603	869
6	79	163	291	474	719	1037
7	91	189	338	550	835	1204
8	104	214	384	625	950	1371
9	116	240	431	701	1065	1537
10	128	266	477	777	1180	1703
11	141	292	523	852	1295	1868
12	153	317	570	928	1410	2035

$k = 3$	$n = 13$	14	15	16	20
	$L_\alpha(k,n) = 165$	178	190	202	251

При $n > 10$ справедлива аппроксимация

$$L^* = \frac{L - \mathbf{M}(L)}{\mathbf{D}(L)},$$

где

$$\mathbf{M}(L) = \frac{nk(k+1)^2}{4}; \quad \mathbf{D}(L) = \frac{n(k^3-k)^2}{144(k-1)}.$$

При $L^* \geqslant u_\alpha$ нулевая гипотеза отклоняется (u_α — α-квантиль стандартного нормального распределения).

Задача 256. *Проверить гипотезу сдвига критерием Пейджа в условиях задачи* 244.

Имеем таблицу:

Номер элемента в выборке	Номер выборки			
	1	2	3	4
1	1	3	4	7
2	2	4	5	8
3	3	5	6	9
4	4	6	7	10
5	5	7	8	11

Заменяя элементы выборки по строкам их рангами в строке, получаем таблицу рангов R_i:

j	i			
	1	2	3	4
1	1	2	3	4
2	1	2	3	4
3	1	2	3	4
4	1	2	3	4
5	1	2	3	4
$\sum$	$R_1 = 5$	$R_2 = 10$	$R_3 = 15$	$R_4 = 20$

Статистика критерия Пейджа равна

$$L = \sum_{i=1}^{k} i \cdot R_i = 1 \cdot 5 + 2 \cdot 10 + 3 \cdot 15 + 4 \cdot 20 = 150.$$

Из табл. 156 для $k = 4$ и $n = 5$ находим $L_{0,95}(4,5) = 137$.
Так как $L = 150 > L_{0,95}(4.5) = 137$, гипотеза сдвига не отклоняется.
Используем теперь аппроксимацию

$$\mathbf{M}(L) = \frac{5 \cdot 4 \cdot 5^2}{4} = 125; \quad \mathbf{D}(L) = \frac{5 \cdot (4^3 - 4)^2}{144 \cdot 3} = 41{,}667;$$

$$L^* = \frac{150 - 125}{\sqrt{41{,}667}} = 3{,}872.$$

Так как $u_{0,95} = 1{,}645$ и $L^* = 3{,}872 > u_{0,95} = 1{,}645$, то и в этом случае гипотеза сдвига не отклоняется.

4.2.1.2.15. Критерий Фридмена–Кендалла–Бэбингтона Смита

Обратимся, как и в предыдущем разделе, к таблице:

Номер элемента в выборке	Номер выборки			
	1	2	...	k
1	x_{11}	x_{21}	...	x_{k1}
2	x_{12}	x_{22}	...	x_{k2}
⋮	⋮	⋮	...	⋮
n	x_{1n}	x_{2n}	...	x_{kn}

Проранжируем, от наибольшего к наименьшему, наблюдения внутри каждой строки. Пусть r_{ij} — ранг члена x_{ij} в совместной ранжировке $x_{1j}, \ldots, x_{kj}$. Обозначим

$$R_i = \sum_{j=1}^{n} r_{ij}; \quad \bar{R}_i = \frac{R_i}{n_i}; \quad \bar{\bar{R}} = \frac{k+1}{2}.$$

Для проверки гипотезы сдвига между параметрами положения выборок равного объема в [467, 468] предложен критерий, основанный на статистике

$$S = \frac{12n}{k(k+1)} \sum_{i=1}^{k} \left(\bar{R}_i - \bar{\bar{R}}\right) = \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1).$$

Гипотеза сдвига отклоняется, если $S < S_\alpha(n,k)$, где $S_\alpha(n,k)$ — критическое значение, приведенное в табл. 157, построенной с применением интерполяции по таблице, приведенной в [18].

В связи с тем, что таблицы распределения рассматриваемого критерия составлены для небольшого диапазона значений, широко применяются различные аппроксимации. Приведем наиболее употребляемые из них.

При $n \geqslant 13$ и $k \geqslant 20$ применима аппроксимация

$$S_\alpha(n,k) = \chi^2_\alpha(k-1).$$

Для других значений n и k используется преобразование

$$F = \frac{(n-1)S}{n(k-1)S}.$$

Таблица 157

Критические значения $S_\alpha(n,k)$ критерия Фридмена–Кендалла–Бэбингтона Смита [18]

n	Доверительная вероятность $\alpha = 0{,}95$			Доверительная вероятность $\alpha = 0{,}90$		
	k			k		
	3	4	5	3	4	5
3	5,82	7,00	8,30	5,42	6,20	7,47
4	6,31	7,20	8,80	5,10	6,00	7,58
5	6,10	7,32	8,96	5,21	6,12	7,61
6	6,33	7,40		4,83	6,20	
7	6,00	7,63		4,71	6,26	
8	6,25	7,50		5,00	6,30	
9	6,00			4,67		
10	6,10			4,90		
11	6,09			4,91		
12	6,08			4,67		
13	6,00			4,77		

Гипотеза сдвига отклоняется, если $F < F_\alpha(f_1, f_2)$, где $F_\alpha(f_1, f_2)$ — α-квантиль распределения Фишера при f_1 и f_2 степенях свободы.

При $7 \leqslant k \leqslant 19$ и $n \geqslant 13$ принимаем $f_1 = k-1$ и $f_2 = (k-1)(n-1)$.

При $k \geqslant 8$ и $7 \leqslant n \leqslant 12$ принимаем $f_1 = k-1$ и

$$f_2 = \frac{L^2}{(n-1)\sum\limits_{1\leqslant i\leqslant k} V_i^2} - (n-1),$$

где

$$V_i = \frac{1}{n-1}\sum_{j=1}^{n}\left(r_{ij} - \bar{R}_i\right); \quad L = (n-1)\sum_{i=1}^{k} V_i; \quad i = 1, \ldots, k.$$

Если f_2 становится дробным, то при использовании таблиц следует применять интерполяцию.

Весьма эффективны аппроксимации, предложенные Иманом и Давенпортом [469]. Для $k \leqslant 7$ и $n \geqslant 8$ статистика Имана–Давенпорта вычисляется по формуле $J = \frac{1}{2}[(k-1)F + S]$.

Гипотеза сдвига отклоняется с достоверностью α, если $J < J_\alpha$, где $f_1 = k-1$, $f_2 = (k-1)(n-1)$ и

$$J_\alpha = \frac{1}{2}\left[(k-1)F_\alpha(f_1, f_2) + \chi^2_\alpha(k-1)\right].$$

Здесь $\chi^2_\alpha(k-1)$ — α-квантиль распределения хи-квадрат с $(k-1)$ степенями свободы.

При $k \geqslant 6$ и $2 \leqslant n \leqslant 6$ и $k = 5$, $n = 6, 7$ вместо J и J_α следует использовать

$$J^* = \frac{1}{2}\left[(k-1)(n-1)F + S\right];$$

$$J^*_\alpha = \frac{1}{2}\left[(k-1)(n-1)F_\alpha(f_1, f_2) + \chi^2_\alpha(k-1)\right].$$

Задача 257. *Для трех выборок случайных величин, приведенных в таблице, проверить гипотезу сдвига критерием Фридмена–Кендалла–Бэбингтона Смита при достоверности* $\alpha = 0{,}95$:

Номер элемента в выборке	Номер выборки			$\bar{x}_j$
	1	2	3	
1	2,1	3,2	4,3	3,20
2	1,8	4,1	2,3	2,73
3	1,7	2,3	3,4	2,47
4	1,8	2,4	3,5	2,57
5	1,9	2,5	3,6	2,67
6	2,4	1,2	3,7	2,43
7	1,7	1,9	3,2	2,27
8	1,6	2,3	2,8	2,23
9	1,5	2,4	2,9	2,26
10	1,7	2,9	3,7	2,83

Строим таблицу рангов r_{ij} по строкам:

j	i		
	1	2	3
1	1	2	3
2	1	3	3
3	1	2	3
4	1	2	3
5	1	2	3
6	2	1	3
7	1	2	3
8	1	2	3
9	1	2	3
10	1	2	3

Имеем

$$R_1 = 11, \quad R_2 = 20, \quad R_3 = 29.$$

Вычисляем статистику критерия

$$S = \frac{12}{10 \cdot 3 \cdot 4} \cdot \sum_{i=1}^{3} R_i^2 - 3 \cdot 10 \cdot 4 = 0{,}1 \cdot \left(11^2 + 20^2 + 29^2\right) - 120 = 16{,}2.$$

Находим из табл. 157 $S_{0,95}(16{,}3) = 6{,}1$.

Так как $S = 16{,}2 > S_{0,95}(16{,}3) = 6{,}1$, гипотеза сдвига не отклоняется.

Продемонстрируем теперь применение аппроксимаций. У нас $k = 3 < 7$ и $n = 10 > 8$, поэтому воспользуемся аппроксимацией J Имана–Давенпорта. Находим

$$F = \frac{(n-1) \cdot S}{n \cdot (k-1) - S} = \frac{9 \cdot 16,2}{10 \cdot 2 - 16,2} = 38,37; \quad J = \frac{1}{2} \cdot (2 \cdot 38,37 + 16,2) = 46,47.$$

Из таблиц находим $F_{0,95}(2; 18) = 3{,}55$ и $\chi^2_{0,95}(2) = 5{,}99$. Следовательно,

$$J_{0,95} = 0{,}5 \cdot (2 \cdot 3{,}55 + 5{,}99) = 6{,}54.$$

Так как $J = 46{,}47 > J_{0,95} = 6{,}54$, гипотеза сдвига не отклоняется.

4.2.1.2.16. Критерий Андерсона–Каннемана–Шэча

Является аналогом критерия Фридмена–Кендалла–Бэбингтона Смита (см. раздел 4.2.1.2.15). Предложен в [470–472]. Для некоторых альтернатив обладает по сравнению с критерием Фридмена большей мощностью. Для пояснения последова-

тельности вычисления статистики критерия воспользуемся обозначениями предыдущего раздела. Так же, как и в критерии Фридмена–Кендалла–Бэбингтона Смита, предварительно ранжируем наблюдения в каждой строке. Затем вычисляем элементы матрицы сопряженности рангов $D = \|D_{li}\|$, где D_{li} — число строк, в которых столбец i получил ранг l $(i, l = 1, 2, \ldots, k)$. Правильность вычислений проверяется условием

$$k\sum_{l=1}^{k} D_{li} = k\sum_{i=1}^{k} D_{li} = \sum_{l=1}^{k}\sum_{i=1}^{k} D_{li} = nk.$$

Статистика критерия вычисляется по формуле

$$A = \frac{k}{n}\sum_{l=1}^{k}\sum_{i=1}^{k}\left(D_{li} - \frac{n}{k}\right)^2 = \frac{k}{n}\sum_{l=1}^{k}\sum_{i=1}^{k} D_{li}^2 - nk.$$

При $n \geqslant 10$ гипотеза сдвига отклоняется, если

$$\frac{k-1}{k}\chi^2_\alpha\left[(k-1)^2\right],$$

где $\chi^2_\alpha(f)$ — α-квантиль распределения χ^2 с f степенями свободы.

Задача 258. *Проверить гипотезу сдвига для данных задачи 257 критерием Андерсона–Каннемана–Шэча.*

Составляем матрицу сопряженности рангов $\|D\|$:

l	i			$\sum D_{li}$
	1	2	3	
1	9	1	0	10
2	1	8	1	10
3	0	1	9	10
$\sum D_{li}$	10	10	10	30

Вычисляем статистику критерия

$$A = \frac{3}{10}\cdot\left(\sum_{l=1}^{3}\sum_{i=1}^{3} D_{li}^2\right) - 10\cdot 3 = 0,3\cdot\left(9^2 + 1^2 + \ldots + 1^2 + 9^2\right) - 30 = 39;$$

$$\frac{k-1}{k}\cdot A = \frac{2}{3}\cdot 39 = 26.$$

Из табл. 55 находим $\chi^2_{0,95}(4) = 9{,}49$. Так как $\frac{k-1}{k}\cdot A = 26 > \chi^2_{0,95}(4) = 9{,}49$, наличие сдвига между параметрами выборок на выбранном уровне значимости $1-\alpha = 0{,}05$ следует считать значимым.

4.2.1.2.17. Критерий со взвешенными ранжировками Даны Квейд

Критерий подобен критерию Фридмена (см. раздел 4.2.1.2.15), но позволяет учесть разброс наблюдений в строках таблицы

j	i			
	1	2	...	k
1	x_{1x}	x_{12}	...	x_{1k}
2	x_{21}	x_{22}	...	x_{2k}
⋮	⋮	⋮	...	⋮
n	x_{n1}	x_{n2}	...	x_{nk}

Критерий предложен в [473, 474]. Ранжирование наблюдений производится внутри строк так же, как и в критериях Фридмена (см. раздел 4.2.1.2.15) и Андерсона (см. раздел 4.2.1.2.16).

Предварительно в качестве „меры доверия" данных выбирается некоторый показатель, например, D_j — дисперсия в j-й строке таблицы. Затем значения дисперсий D_j ранжируются рангами q_j от наименьшего к наибольшему ($j = 1, 2, \ldots, n$). Вычисляются величина

$$S = \sum_{i=1}^{k}\left(\sum_{j=1}^{n} q_j r_{ij}\right)^2$$

и статистика критерия

$$W = \frac{72S}{k(k+1)n(n+1)(2n+1)} - \frac{9(k+1)(n+1)n}{2(2n+1)}.$$

При $n \geqslant 5$ справедлива аппроксимация

$$W^* = \frac{[W-(k-1)]c}{d} + \delta,$$

где

$$c = 1 - 6\frac{3n^2+3n-1}{5n(n+1)(2n+1)}; \quad \delta = \frac{(k-1)c^3}{d}; \quad d = 3c - 2 + \frac{72(3n^4+6n^3-3n+1)}{7n^2(n+1)^2(2n+1)^2}.$$

Гипотеза сдвига отклоняется с достоверностью α, если $W^* > \chi^2_\alpha(\delta)$, где $\chi^2_\alpha(\delta)$ — α-квантиль распределения хи-квадрат с $f = \delta$ степенями свободы.

В среднем критерий Д. Квейд мощнее критерия Фридмена по отношению к большинству альтернатив [475].

Задача 259. *Проверить гипотезу сдвига для данных задачи 257 критерием Даны Квейд.*

Подсчитываем дисперсии D_j по формуле

$$D_j = \frac{1}{k-1}\sum_{i=1}^{k}(x_{ij} - \bar{x}_i)^2, \quad \text{где} \quad \bar{x}_i = \frac{1}{k}\sum_{i=1}^{k} x_{ij}.$$

Имеем таблицу рангов r_{ij} по строкам и дисперсий D_j по строкам, ранжированных внутри столбца от 1 до n:

j	i			D_j	q_j
	1	2	3		
1	1	2	3	1,21	8
2	1	3	2	1,46	9
3	1	2	3	0,74	4
4	1	2	3	0,74	5
5	1	2	3	0,74	6
6	2	1	3	1,56	10
7	1	2	3	0,66	3
8	1	2	3	0,36	1
9	1	2	3	0,50	2
10	1	2	3	1,01	7

Вычисляем

$$S = \sum_{i=1}^{3}\left(\sum q_j \cdot r_{ij}\right)^2 = \left(\sum_{j=1}^{10} q_j \cdot r_{1j}\right)^2 + \left(\sum_{j=1}^{10} q_j \cdot r_{2j}\right)^2 + \left(\sum_{j=1}^{10} q_j \cdot r_{3j}\right)^2 =$$
$$= (1\cdot 8 + 1\cdot 9 + 1\cdot 4 + \ldots + 1\cdot 7) + \ldots + (3\cdot 8 + 3\cdot 9 + 3\cdot 4 + \ldots + 3\cdot 2 + 3\cdot 7)^2 = 40442;$$

$$W = \frac{72 \cdot 40442}{3 \cdot 4 \cdot 10 \cdot 11 \cdot 21} - \frac{9 \cdot 4 \cdot 11 \cdot 10}{2 \cdot 21} = 10{,}758.$$

Далее вычисляем

$$c = 1 - 6 \cdot \frac{3 \cdot 100 + 3 \cdot 10 - 1}{5 \cdot 10 \cdot 11 \cdot 21} = 0{,}829; \quad \delta = \frac{2 \cdot 0{,}829^2}{0{,}556^2} = 4{,}446;$$

$$d = 3 \cdot 0{,}829 - 2 + \frac{72 \cdot \left(3 \cdot 10^4 + 6 \cdot 10^3 - 3 \cdot 10 + 1\right)}{7 \cdot 10^2 \cdot 11^2 \cdot 21^2} = 0{,}556.$$

Тогда $W^* = \dfrac{(10{,}758 - 2) \cdot 0{,}829}{0{,}556} + 4{,}446 = 17{,}504.$

Из табл. 55 находим $\chi^2_{0,95}(3{,}686) \approx 8{,}96$.

Так как $W^* = 17{,}504 > \chi^2_{0,95}(3{,}686) = 8{,}96$, гипотеза сдвига не отклоняется.

4.2.1.2.18. Критерий Кендалла–Эренберга

Является аналогом критерия Фридмена–Кендалла–Бэбингтона Смита (см. раздел 4.2.1.2.15), превосходя его по мощности [477]. Предварительно наблюдения ранжируются по строкам, как и в предыдущих критериях. Обозначим через c_{lj} число тех случаев, когда ранг l-го объекта больше ранга j-го объекта $(1 \leqslant l,\ j < k)$ и c_{jl} — число тех случаев, когда ранг l-го элемента меньше ранга j-го.

Предварительно вычисляются элементы матрицы $\|c_{lj}\|$ (при этом $c_{jj} = 0$). Тогда статистика критерия имеет вид

$$K = \sum_{1 \leqslant l < j \leqslant k} c_{lj}(n - c_{lj}) = n \sum_{1 \leqslant l < j \leqslant k} c_{lj} - \sum_{1 \leqslant l < j \leqslant k} c_{lj}^2.$$

Гипотеза не отклоняется, если $K > K_\alpha(k, n)$, где $K_\alpha(k, n)$ — критическое значение, приведенное для некоторых n и k в табл. 158.

Таблица 158

Критические значения $K_\alpha(k, n)$ критерия Кендалла–Эренберга для доверительной вероятности $\alpha = 0{,}95$ [119]

k	n	$K_{0,95}$	k	n	$K_{0,95}$	k	n	$K_{0,95}$
3	4	3	3	9	42	4	6	34
3	5	18	3	10	54	5	3	8
3	6	14	4	3	4	5	4	20
3	7	22	4	4	10			
3	8	31	4	5	20			

Для выборок большого объема применяется следующее приближение, основанное на статистике U^*,

$$U^* = aU + f,$$

где

$$a = \frac{6(2k+5)kn(k-1)(n-1)}{4(n-2)(2k^2+6k+7)}; \quad U = 1 - \frac{8k}{kn(k-1)(n-1)};$$

$$f = \frac{2(3k+5)^3 kn(k-1)(n-1)}{4(n-2)^2(2k^2+6k+7)^2}.$$

Гипотеза сдвига отклоняется, если $U^* < \chi^2_\alpha(f)$, где $\chi^2_\alpha(f)$ — α-квантиль χ^2-распределения с f степенями свободы. Следует помнить, что в этом критерии, как и во всех предыдущих, для одинаковых (связанных) наблюдений применяются средние ранги (в нашем случае связи $c_{lj} = 1$ и $c_{jl} = 0$ заменяются на $c_{lj} = c_{jl} = 1/2$).

Задача 260. *Проверить гипотезу сдвига для данных задачи 257 критерием Кендалла–Эренберга.*

Восстановим таблицу рангов r_{ij} из задачи 257:

j	i		
	1	2	3
1	1	2	3
2	1	3	2
3	1	2	3
4	1	2	3
5	1	2	3
6	2	1	3
7	1	2	3
8	1	2	3
9	1	2	3
10	1	2	3

Вычисляем элементы матрицы $||c_{lj}||$:

l	j		
	1	2	3
1	*	1	0
2	9	*	1
3	10	9	*

Из элементов второго столбца ($i = 2$) только один элемент имеет ранг, меньший, чем элементы первого столбца. Следовательно, $c_{12} = 1$. Аналогичным образом заполняются все остальные элементы матрицы $||c_{lj}||$.

Вычисляем

$$K = c_{12} \cdot c_{21} + c_{13} \cdot c_{31} + c_{23} \cdot c_{32} = c_{12} \cdot (n - c_{12}) + c_{13} \cdot (n - c_{13}) + c_{23} \cdot (n - c_{23}) = \\ = 1 \cdot (10 - 1) + 0 \cdot (10 - 0) + 9 \cdot (10 - 9) = 18.$$

Из табл. 158 для $k = 3$ и $n = 7$ находим $K_{0,95} = 22$.

Так как $K = 18 < K_{0,95} = 22$, гипотеза сдвига не отклоняется.

Вычислим теперь приближенный критерий. Находим

$$a = \frac{6 \cdot 11 \cdot 3 \cdot 10 \cdot 2 \cdot 9}{4 \cdot 8 \cdot (2 \cdot 9 + 6 \cdot 3 + 7)} = 25{,}9; \quad U = 1 - \frac{8 \cdot 18}{3 \cdot 10 \cdot 2 \cdot 9} = 0{,}7333;$$

$$f = \frac{2 \cdot 11^3 \cdot 3 \cdot 10 \cdot 2 \cdot 9}{4 \cdot 8^2 \cdot (2 \cdot 3^2 + 6 \cdot 3 + 7)^2} = 3{,}0369.$$

Далее $U^* = 25{,}90 \cdot 0{,}733 + 3{,}037 = 22{,}02$. Из табл. 55 имеем $\chi^2_{0,95}(3) = 7{,}81$.

Так как $U^* = 22{,}02 > \chi^2_{0,95}(3) = 7{,}81$, гипотеза сдвига не отклоняется.

4.2.1.2.19. Критерий Ходжеса–Лемана–Сена

Аналог критерия Фридмена--Кендалла--Бэбингтона Смита (см. раздел 4.2.1.2.15), но обладает по сравнению с ним большей эффективностью. Статистика критерия строится следующим образом. Находится среднее значение в j-й строке $\bar{x}_j = \dfrac{1}{k} \sum\limits_{i=1}^{k} x_{ij}$, и определяются выровненные наблюдения $y_{ij} = x_{ij} - \bar{x}_j$. Затем все $(n-k)$ выровненных наблюдений y_{ij} ранжируются, что приводит к последовательности рангов R_{ij}.

Далее, пусть

$$z_{N,\nu}^{(i)} = \begin{cases} 1, & \text{если } \nu\text{-е наблюдение среди наблюдений } y_{ij} \text{ относится к } i\text{-й строке;} \\ 0 & \text{в противном случае } (i = 1, 2, \ldots, 2k;\ \nu = 1, 2, \ldots, N). \end{cases}$$

Введем обозначения

$$T_{N,i} = \frac{1}{n} \sum_{\nu=1}^{N} z_{N,\nu}^{(i)} \nu, \quad (i = 1, \ldots, k); \quad \bar{R}_j = \frac{1}{k} \sum_{ij=1}^{k} R_{ij};$$

$$M = \frac{1}{n(k-1)} \left[\frac{N(N-1)(2N+1)}{6} - k \sum_{j=1}^{k} \bar{R}_j^2 \right].$$

Тогда статистика критерия имеет вид $S = \frac{n}{M} \left[\sum_{i=1}^{k} T_{N,i}^2 - k\left(\frac{N+1}{2}\right)^2 \right]$.

При $n \geqslant 10$ гипотеза сдвига отклоняется, если $S < \chi_\alpha^2(k-1)$, где χ_α^2 — α-квантиль распределения хи-квадрат с $f = k - 1$ степенями свободы.

Следует помнить, что в этом критерии для равных наблюдений не применяются средние ранги, они просто распределяются между равными величинами случайным образом.

Задача 261. *Проверить гипотезу сдвига для данных задачи* 257 *критерием Ходжеса–Лемана–Сена.*

Иллюстрируем вычисления по шагам. Вычисляем величины y_{ij}:

j	i			j	i		
	1	2	3		1	2	3
1	−1,10	0	1,10	6	−0,03	−1,23	1,27
2	9,93	1,37	−0,43	7	−0,57	−0,37	0,93
3	−0,77	−0,17	0,93	8	−0,73	−0,03	0,47
4	−0,77	−0,17	0,93	9	−0,77	0,13	0,63
5	−0,77	−0,17	0,93	10	−1,07	0,13	0,93

Ранги R_{ij} величин y_{ij} составляют следующую таблицу (равным значениям y_{ij} случайным образом присваиваются последовательные ранги без усреднения):

j	i			j	i		
	1	2	3		1	2	3
1	2	18	28	6	16	1	29
2	4	30	11	7	10	12	27
3	5	13	24	8	8	17	21
4	6	14	25	9	9	19	22
5	7	15	26	10	3	20	23

Вычислим теперь значения $z_{N,\nu}^{(i)}$, $i = 1, 2, 3$ и $\nu = 1, \ldots, 30$. Всего нужно вычислить $nN = 10 \cdot 30 = 300$ значений величин $\left\{z_{N,\nu}^{(i)}\right\}$. Затем необходимо найти те значения $z_{N,\nu}^{(i)}$ $(i = 1, 2, 3)$, которые равны 1. Результаты сведем в следующую таблицу значений для различных ν и i при $N = 30$:

ν	i			ν	i			ν	i		
	1	2	3		1	2	3		1	2	3
1		1		11			1	21			1
2	1			12		1		22			1
3	1			13		1		23			1
4	1			14		1		24			1
5	1			15		1		25			1
6	1			16	1			26			1
7	1			17		1		27			1
8	1			18		1		28			1
9	1			19		1		29			1
10	1			20		1		30		1	

Из таблицы имеем

$$T_{N,1} = \frac{1}{10} \cdot \sum z_{N,\nu}^{(1)} \cdot \nu = \frac{2+3+4+5+6+7+8+9+10+16}{10} = 710;$$

$$T_{N,2} = \frac{1}{10} \cdot \sum_{i=1}^{10} z_{N,\nu}^{(2)} \cdot \nu = \frac{12+13+14+15+17+18+19+20+30}{10} = 15{,}8$$

$$T_{N,3} = \frac{1}{10} \cdot \sum_{i=1}^{10} z_{N,\nu}^{(3)} \cdot \nu = \frac{11+21+22+23+24+25+26+27+28+29}{10} = 23{,}6.$$

Далее, вычисляя $\bar{R}_j = \frac{1}{3} \cdot \sum_{i=1}^{3} R_{ij}$, имеем $\bar{R}_1 = 16$; $\bar{R}_2 = 15$; $\bar{R}_3 = 14$; $\bar{R}_4 = 16$; $\bar{R}_6 = 15{,}33$; $\bar{R}_7 = 16{,}33$.

Тогда

$$\sum \bar{R}_j^2 = 2407{,}5845; \quad M = 111{,}6123; \quad S = 12{,}366.$$

Имеем из табл. 55 $\chi^2_{0,95}(2) = 5{,}99$. Так как $S = 12{,}366 > \chi^2_{0,95}(2) = 5{,}99$, гипотеза сдвига не отклоняется.

4.2.2. Непараметрические критерии масштаба

Непараметрические ранговые критерии сравнения параметров масштаба, как правило, строятся на базе соответствующих критериев сдвига изменением либо статистики критерия, либо правил присвоения рангов наблюдениям. При дальнейшем изложении достаточно широкой гаммы известных критериев масштаба мы будем специально обращать внимание читателя на это обстоятельство. Напомним, что критерии масштаба преследуют цель выявить возможные различия в мерах разброса (изменчивости) наблюдений в двух или более выборках.

4.2.2.1. Сравнение параметров масштаба двух совокупностей

4.2.2.1.1. Критерий Ансари–Бредли

Является масштабным аналогом критерия Вилкоксона (см. раздел 4.2.1.1.2.2). Сравниваются две выборки x_i и y_i, объемами m и n соответственно. Пусть R_i — ранги элементов одной из выборок (предположим, x) в общем упорядоченном по возрастанию ряду. Статистикой критерия Ансари–Бредли является

$$S = \sum_{i=1}^{m} \left\{ \frac{m+n+1}{2} - \left| R_i - \frac{m+n+1}{2} \right| \right\}.$$

Вычисление статистики критерия может быть выполнено и другим, более простым методом. Поставим элементам упорядоченной по возрастанию выборки $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_{m+n}$ объема $m+n$ в соответствие ранги по следующему правилу

$$R(x_i) = \begin{cases} m+n-i+1, & \text{если } i > \dfrac{m+n+1}{2}; \\ i, & \text{если } i \leqslant \dfrac{m+n+1}{2}. \end{cases}$$

Тогда статистика критерия равна

$$S = \sum_{i=1}^{m} R(x_i),$$

т. е. она определяется суммой специальным образом назначенных рангов одной выборки.

Легко видеть, что при четном $(m+n)$ последовательность таких рангов имеет вид

$$1, 2, 3, \ldots, \frac{m+n}{2}, \frac{m+n}{2}, \ldots, 3, 2, 1,$$

а при нечетном $(m+n)$ —

$$1, 2, 3, \ldots, \frac{m+n-1}{2}, \frac{m+n+1}{2}, \ldots, 3, 2, 1.$$

Гипотеза равенства параметров масштаба не отклоняется с достоверностью α, если $S_1(\alpha) < S < S_2(\alpha)$, где $S_1(\alpha)$, $S_2(\alpha)$ — критические значения, приведенные в табл. 159.

При $m, n > 10$ можно использовать асимптотическую нормальность распределения величины

$$S^* = \frac{S - \mathbf{M}(S)}{\sqrt{\mathbf{D}(S)}},$$

где

$$\mathbf{M}(S) = \begin{cases} \dfrac{m(m+n+2)}{4} & \text{при } m+n = 2k; \\ \dfrac{m(m+n+1)^2}{4(m+n)} & \text{при } m+n = 2k-1; \end{cases}$$

$$\mathbf{D}(S) = \begin{cases} \dfrac{mn(m+n-2)(m+n+2)}{48(m+n-1)} & \text{при } m+n = 2k; \\ \dfrac{mn(m+n+1)[(m+n)^2+3]}{48(m+n)^2} & \text{при } m+n = 2k-1. \end{cases}$$

Нулевая гипотеза равенства параметров масштаба в двух выборках принимается с достоверностью α, если

$$|S^*| < u_{\frac{1+\alpha}{2}}.$$

Эффективность критерия по сравнению с F-критерием в случае нормального распределения равна $\dfrac{6}{\pi^2} \approx 0{,}61$.

Таблица 159

Критические значения $S_1(\alpha)$ и $S_2(\alpha)$ статистики Ансари–Бредли (α — доверительная вероятность) [479]

m	n	α				m	n	α			
		0,90		0,95				0,90		0,95	
		S_1	S_2	S_1	S_2			S_1	S_2	S_1	S_2
2	8	2	10	2	10	5	5	10	20	10	20
	9	2	11	2	11		6	1	21	10	23
	10	2	12	2	12		7	11	24	11	24
	11	2	13	2	13		8	12	26	11	26
	12	2	14	2	14		9	13	27	12	28
	13	3	14	2	15		10	14	29	12	30
	14	3	15	2	16		11	14	31	13	32
	15	3	16	2	17		12	15	33	14	34
	16	3	17	2	17		13	16	34	14	36
	17	3	18	2	19		14	16	36	15	38
	18	3	19	2	19		15	17	38	15	40
3	6	4	13	4	13	6	6	15	27	14	28
	7	5	13	4	14		7	16	29	15	30
	8	5	15	4	16		8	17	31	16	32
	9	5	16	4	17		9	18	34	16	35
	10	5	17	5	18		10	18	36	17	37
	11	6	18	5	19		11	19	38	18	40
	12	6	20	5	21		12	20	40	19	41
	13	6	21	5	22		13	21	42	19	44
	14	7	22	6	23		14	22	44	20	46
	15	7	23	6	24	7	7	21	35	19	37
	16	7	24	6	25		8	22	38	20	39
	17	8	25	6	26		9	23	40	21	42
4	5	7	14	6	16		10	24	43	22	44
	6	7	17	7	17		11	25	45	23	47
	7	8	19	7	19	8	8	26	45	26	46
	8	8	20	7	21		9	29	48	27	49
	9	9	21	8	22		10	30	50	28	52
	10	9	23	8	24		11	31	53	29	55
	11	10	24	9	26		12	32	56	30	58
	12	10	26	9	27	9	9	35	55	33	57
	13	11	27	9	29		10	36	58	34	58
	14	11	29	10	30		11	38	61	36	63
	15	12	30	10	32	10	10	43	67	41	69
	16	12	32	11	33						

Задача 262. *Имеются две выборки случайных величин:*

$$(m = 8)\ x_{1j}:\quad 1{,}2;\ \ 3{,}4;\ \ 6{,}2;\ \ 8{,}1;\ \ 10{,}2;\ \ 11{,}3;\ \ 13{,}0;\ \ 15{,}9;$$

$$(n = 10)\ x_{2j}:\quad 0{,}8;\ \ 2{,}4;\ \ 4{,}2;\ \ 5{,}1;\ \ 6{,}8;\ \ 11{,}4;\ \ 13{,}8;\ \ 20{,}1;\ \ 24{,}2;\ \ 26{,}7.$$

Проверить гипотезу равенства параметров масштаба в выборках критерием Ансари–Бредли при достоверности $\alpha = 0{,}95$.

Упорядочиваем объединенную выборку по возрастанию (i — номер выборки, R_i — ранг случайной величины в объединенной выборке). Результаты сведем в таблицу:

i	x_{ij}	R_{ij}	i	x_{ij}	R_{ij}	i	x_{ij}	R_{ij}
2	0,8	1	1	6,2	7	1	13,0	13
1	1,2	2	2	6,8	8	2	13,8	14
2	2,4	3	1	8,1	9	1	15,9	15
1	3,4	4	1	10,2	10	2	20,1	16
2	4,2	5	1	11,3	11	2	24,2	17
2	5,1	6	2	11,4	12	2	26,7	18

Далее находим

$$S = \sum_{i=1}^{8} \left\{ \frac{8+10+1}{2} - \left| R_i - \frac{8+10+1}{2} \right| \right\} = \sum_{i=1}^{8} \{9{,}5 - |R_i - 9{,}5|\} =$$
$$= 76 - (|2-9{,}5| + |4-9{,}5| + |7-9{,}5| + \ldots + |13-9{,}5| + |15-9{,}5|) = 49.$$

Если бы использовался иной метод присвоения рангов, то имели бы следующую последовательность рангов:

i:	2	1	2	1	2	2	1	2	1	1	1	2	1	2	1	2	2	2
R_i:	1	2	3	4	5	6	7	8	9	9	8	7	6	5	4	3	2	1

Сумма таких рангов первой выборки равна $S = 2+4+7+9+9+8+6+4 = 49$, что естественно совпадает с ранее полученным результатом.

Теперь из табл. 159 для $m = 8$, $n = 10$ и $\alpha = 0{,}95$ находим $S_1 = 28$ и $S_2 = 52$. Так как $S_1 = 28 < S = 49 < S_2 = 51$, нулевая гипотеза равенства параметров масштаба в выборках не отклоняется.

Используем теперь нормальную аппроксимацию. Так как $(m+n)$ — четное, имеем

$$\mathbf{M}(S) = \frac{8 \cdot 20}{4} = 40; \quad \mathbf{D}(S) = \frac{8 \cdot 10 \cdot 16 \cdot 20}{48 \cdot 17} = 31,3725 \quad (\sqrt{\mathbf{D}(S)} = 5,601);$$
$$S^* = \frac{49-40}{5,601} \cdot 1,606.$$

Так как $S^* = 1{,}606 < u_{(1+\alpha)/2} = u_{0{,}975} = 1{,}96$, нулевая гипотеза и в этом случае не отклоняется.

4.2.2.1.2. Критерий Сижела–Тьюки

Сижел и Тьюки [480] предложили преобразование критериев сдвига в критерии масштаба. Суть их способа сводится к преобразованию первичной упорядоченной объединенной выборки. Пусть $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ — первичная объединенная выборка. Из нее получаем новую последовательность вида

$$x_1, x_n, x_{n-1}, x_2, x_3, x_{n-2}, x_{n-3}, x_4, x_5, \ldots$$

(т. е. оставшийся ряд „переворачивается“ каждый раз после приписывания рангов паре крайных значений).

Далее проверка гипотезы о разности параметров масштаба в двух выборках аналогична проверке гипотезы сдвига в новой последовательности с описанным правилом нумерации рангов.

Если использовать в качестве критерия проверки нулевой гипотезы сумму рангов ($R = \sum_{i=1}^{m} R_i$) элементов выборки меньшего объема в такой последовательности, то нулевая гипотеза принимается, если $R_1(\alpha) < R < R_2(\alpha)$, где $R_1(\alpha)$ и $R_2(\alpha)$ — критические значения, которые могут быть получены с помощью табл. 143 критерия Манна–Уитни (см. раздел 4.2.1.1.2.2). Для этого необходимо из табл. 143 найти

U_1 и U_2 для заданных α, m и n и затем вычислить

$$R_1(\alpha) = nm + \frac{m(m+1)}{2} - U_2; \quad R_2(\alpha) = nm + \frac{m(m+1)}{2} - U.$$

Здесь m — объем меньшей выборки.

При $n, m > 10$ справедлива аппроксимация

$$W = \frac{R - \dfrac{m(n+m+1)}{2}}{\sqrt{\dfrac{nm(n+m+1)}{2}}} \to N(0,1).$$

Если $|W| > u_{\frac{1+\alpha}{2}}$, нулевая гипотеза равенства параметров масштаба принимается с достоверностью α.

Задача 263. *Проверить гипотезу равенства параметров масштаба критерием Сижела–Тьюки для данных задачи* 262.

Строим упорядоченный ряд с нумерацией рангов в соответствии с правилом критерия Сижела–Тьюки. Результаты сведем в таблицу (i — номер выборки; R_i — ранг члена объединенной выборки, упорядоченной по правилу Сижела–Тьюки).

i	x_i	R_i	i	x_i	R_i	i	x_i	R_i
2	0,8	1	1	15,9	7	1	6,2	13
2	26,7	2	1	3,4	8	2	11,4	14
2	24,2	3	2	4,2	9	1	11,3	15
1	1,2	4	2	13,8	10	2	6,8	16
2	2,4	5	1	13,0	11	1	8,1	17
2	20,1	6	2	5,1	12	1	10,2	18

Для первой выборки ($i = 1$) из таблицы имеем

$$R = \sum_{i=1}^{8} R_i = 4 + 7 + 8 + 11 + 13 + 15 + 17 + 18 = 93.$$

Из табл. 143 для $m = 8$, $n = 10$ и $\alpha = 0{,}95$ находим $U_1 = 17$ и $U_2 = 63$, откуда $R_1(\alpha) = 80 + \frac{8 \cdot 9}{2} - 63 = 53$ и $R_2(\alpha) = 80 + \frac{8 \cdot 9}{2} - 17 = 99$.

Так как $R_1(\alpha) = 53 < R = 93 < R_2(\alpha)$, нулевая гипотеза принимается.

Используем теперь аппроксимацию

$$W = \frac{93 - \dfrac{8 \cdot 19}{2}}{\sqrt{\dfrac{10 \cdot 8 \cdot 19}{12}}} = 1{,}51.$$

Так как $W = 1{,}51 < u_{0,975} = 1{,}96$, нулевая гипотеза и в этом случае не отклоняется.

4.2.2.1.3. Критерий Кейпена

Предложен в [481] и является масштабным аналогом критерия Фишера–Йэйтса–Терри–Гёфдинга (см. раздел 4.2.1.1.2.3).

Если R_i — ранг i-го элемента меньшей по объему выборки $x_1, x_2, \ldots, x_m$ в общем упорядоченном ряду $(m + n)$ значений объединенной выборки, то статистика критерия может быть записана в виде

$$K = \sum_{i=1}^{m} a_{m+n}(R_i), \quad m \leqslant n,$$

где $a_{m+n}(i)$ — математическое ожидание квадрата i-й порядковой статистики в выборке объема $(m+n)$ из стандартного нормального распределения (значения $a_N(i)$, называемые метками критерия, приведены в табл. 160).

Таблица 160

Метки $a_N(i)$ критерия Кейпена (все значения умножены на 100) [482]

i	N														
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	202	222	240	256	271	285	298	310	321	331	341	351	360	368	376
2	69	83	97	109	122	133	144	155	164	174	182	191	199	207	214
3	29	34	42	51	61	70	79	89	96	104	112	119	126	133	140
4	29	21	21	25	30	36	43	50	57	63	70	77	83	89	95
5	69	34	21	17	17	19	23	27	32	38	43	49	54	60	65
6	202	83	42	25	17	14	14	15	18	22	26	30	34	39	44
7		222	97	51	30	19	14	12	12	13	15	18	21	25	28
8			240	109	61	36	23	15	12	10	10	11	13	15	18
9				256	122	70	43	27	18	13	10	9	9	10	11
10					271	133	79	50	32	22	15	11	9	8	8
11						285	144	87	57	38	26	18	13	10	8
12							298	155	96	63	43	30	21	15	11
13								310	164	104	70	49	34	25	18
14									321	174	112	77	54	39	28
15										331	182	119	83	60	44
16											341	191	126	89	65
17												351	199	133	95
18													360	207	140
19														368	214
20															376

Нулевая гипотеза отклоняется, если

$$K_1(\alpha) < K < K_2(\alpha),$$

где $K_1(\alpha)$ и $K_2(\alpha)$ — критические значения, приведенные в табл. 161.

При $m, n > 10$ справедливо приближение

$$K^* = \frac{K - \mathbf{M}(K)}{\sqrt{\mathbf{D}(K)}} \to N(0,1),$$

где

$$\mathbf{M}(K) = m; \quad \mathbf{D}(K) = \frac{mn}{(m+n)(m+n-1)} \sum_{i=1}^{m+n} a_{m+n}^2(i) - \frac{mn}{m+n-1}.$$

При $|K^*| < u_{\frac{1+\alpha}{2}}$ нулевая гипотеза принимается с достоверностью α.

Задача 264. *Проверить гипотезу равенства параметров масштаба критерием Кейпена для данных задачи* 262.

Ранги R_i первой выборки (объема $m = 8$) в общей упорядоченной последовательности равны: 2, 4, 7, 9, 10, 11, 13, 15.

Из табл. 160 находим ряд $a_N(R_i)$:

$$a_{18}(2) = 199; \quad a_{18}(4) = 83; \quad a_{18}(7) = 21; \quad a_{18}(9) = 9;$$
$$a_{18}(10) = 9; \quad a_{18}(11) = 13; \quad a_{18}(13) = 34; \quad a_{18}(15) = 83.$$

Таблица 161

Критические значения $K_1(\alpha)$ и $K_2(\alpha)$ статистики Кейпена
(все значения умножены на 100, α — доверительная вероятность) [482]

n	m	α				n	m	α			
		0,90		0,95				0,90		0,95	
		K_1	K_2	K_1	K_2			K_1	K_2	K_1	K_2
6	2	42	80	42	80	12	2	30	485	24	642
5	3	84	577	84	577	11	3	76	581	56	660
4	4	181	619	126	674	10	4	133	717	113	795
6	3	93	563	67	621	9	5	201	842	170	895
5	4	176	672	144	672	8	6	276	946	233	991
8	2	34	542	34	542	7	7	366	1034	313	1087
7	3	95	572	77	603	13	2	32	505	26	662
6	4	156	694	125	725	12	3	70	609	57	679
5	5	247	755	199	803	11	4	122	727	102	804
9	2	78	570	33	570	10	5	196	851	158	912
8	3	86	589	69	640	9	6	270	959	228	1016
7	4	139	710	122	739	8	7	357	1054	308	1105
6	5	226	792	192	843	14	2	30	523	25	523
10	2	37	442	28	596	13	3	68	593	51	697
9	3	80	586	71	639	12	4	122	735	99	809
8	4	139	698	123	763	11	5	190	860	159	929
7	5	211	826	173	862	10	6	264	970	221	1030
6	6	297	905	261	941	9	7	345	1072	300	1126
11	2	30	465	27	620	8	8	431	1167	381	1217
10	3	77	620	57	647	14	3	69	619	52	711
9	4	139	719	107	790	13	4	122	743	99	828
8	5	203	829	171	877	11	6	185	863	154	947
7	6	286	927	243	969	10	7	335	1090	292	1158
9	8	423	1186	369	1246	10	8	412	1203	360	1266
11	7	329	1101	280	1170	9	9	501	1297	442	1356

Тогда $K = \sum_{i=1}^{18} a_{18}(i) = 199 + 83 + 21 + 9 + 9 + 13 + 34 + 83 = 451$.

Из табл. 161 для $\alpha = 0{,}95$, $m = 8$ и $n = 10$ находим $K_1 = 360$ и $K_2 = 1266$.

Так как $K_1 = 360 < K = 451 < K_2 = 1266$, нулевая гипотеза равенства параметров масштаба в выборках не отклоняется.

Покажем теперь использование аппроксимации. Необходимо иметь в виду, что в табл. 160 и 161 приведены значения $a_N(i)$ и $K_1(K_2)$, в 100 раз превосходящие реальные величины. Это не сказывается на проверке критерия, однако должно быть учтено при использовании аппроксимации (т. е. при вычислении величины K^* нужно использовать значения K и $a_{n+m}(i)$, уменьшенные в 100 раз). Имеем $\mathbf{M}(K) = 8$ и (используя табл. 161)

$$\sum_{i=1}^{18} a_{18}^2(i) = 3{,}6^2 + 1{,}99^2 + 1{,}26^2 + 0{,}83^2 + 0{,}54^2 + 0{,}34^2 + 0{,}21^2 + 0{,}13^2 + 0{,}09^2 + 0{,}09^2 + \\ + 0{,}13^2 + 0{,}21^2 + 0{,}34^2 + 0{,}54^2 + 0{,}83^2 + 1{,}26^2 + 1{,}99^2 + 3{,}6^2 = 39{,}3458.$$

Далее получаем

$$\mathbf{D}(K) = \frac{8 \cdot 10}{18 \cdot 17} \cdot 39{,}3458 - \frac{8 \cdot 10}{17} 5{,}5806 \quad (\sqrt{\mathbf{D}(K)} = 2{,}362); \quad |K^*| = \frac{|4{,}51 - 8|}{2{,}362} = 1{,}478.$$

Так как $|K^*| = 1{,}478 < u_{\frac{1+\alpha}{2}} = u_{0{,}975} = 1{,}96$, нулевая гипотеза не отклоняется и этим критерием.

4.2.2.1.4. Критерий Клотца

Критерий, предложенный Клотцем [483], является масштабным аналогом критерия Ван дер Вардена (см. раздел 4.2.1.2.6). Его статистика в принятых ранее обозначениях имеет вид $L = \sum\limits_{i=1}^{m} u^2_{\frac{R_i}{m+n+1}}$, где u_γ — γ-квантиль стандартного нормального распределения; $m \leqslant n$ — объемы сравниваемых выборок; R_i — ранги, полученные элементами первой выборки (объема m) в общем упорядоченном по возрастанию ряду. Напомним (см. аппроксимацию 15 в разделе 1.1.1), что можно использовать аппроксимацию

$$u_\gamma \approx 4{,}91\left[\gamma^{0,14} - (1-\gamma)^{0,14}\right].$$

Метки критерия Клотца $u^2_{\frac{i}{N}}$ приведены в табл. 162.

Таблица 162

Метки $u^2_{\frac{i}{N}}$ критерия Клотца (все значения умножены на 100) [482]

i	N														
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	114	132	149	164	178	191	203	215	225	235	245	254	262	271	278
2	32	46	58	71	83	94	104	114	123	132	141	149	157	164	171
3	3	10	19	28	37	46	54	63	71	79	86	94	101	107	114
4	3	0	2	6	12	19	25	32	39	46	52	58	65	71	77
5	32	10	2	0	1	4	9	13	19	24	29	35	40	46	51
6	114	46	19	6	1	0	1	3	6	10	14	19	23	28	32
7		132	58	28	12	4	1	0	1	2	5	8	11	15	19
8			149	71	37	19	9	3	1	0	1	2	4	6	9
9				164	83	46	25	13	6	2	1	0	0	2	1
10					178	94	54	32	19	10	5	2	0	0	0
11						191	104	63	39	24	14	0	4	2	0
12							203	114	71	46	29	19	11	6	1
13								215	123	79	52	35	23	15	1
14									225	132	86	58	40	28	9
15										235	141	94	65	46	19
16											245	149	101	71	32
17												254	157	107	51
18													262	164	77
19														271	114
20															171
															278

Гипотеза равенства параметров масштаба двух выборок принимается с достоверностью α, если $L_1(\alpha) < L < L_2(\alpha)$, где $L_1(\alpha)$ и $L_2(\alpha)$ — критические значения, приведенные в табл. 163.

При $m, n > 10$ распределение статистик L удовлетворительно аппроксимируется нормальным со средним $\mathbf{M}(L)$ и дисперсией $\mathbf{D}(L)$, где

$$\mathbf{M}(L) = \frac{m}{m+n}\sum_{i=1}^{m+n} u^2_{\frac{i}{m+n+1}};$$

$$\mathbf{D}(L) = \frac{mn}{(m+n)(m+n-1)}\sum_{i=1}^{m+n} u^4_{\frac{i}{m+n+1}} - \frac{n}{m(m+n-1)}\left[\frac{m}{m+n}\sum_{i=1}^{m+n} u^2_{\frac{i}{m+n+1}}\right].$$

Таблица 163

Критические значения $L_1(\alpha)$ и $L_2(\alpha)$ статистики Клотца
(все значения умножены на 100, α — доверительная вероятность) [482]

m	n	α				m	n	α			
		0,90		0,95				0,90		0,95	
		L_1	L_2	L_1	L_2			L_1	L_2	L_1	L_2
2	6	4	298	4	298	3	14	37	442	23	508
3	5	23	356	23	356	4	13	75	552	56	602
4	4	81	375	42	414	5	12	122	640	95	694
3	6	34	356	12	399	6	11	177	719	141	772
4	5	83	427	62	427	7	10	234	797	198	847
2	8	2	356	2	356	8	9	298	870	256	919
3	7	39	368	25	393	3	15	34	459	23	524
4	6	76	451	51	476	4	14	74	560	54	616
5	5	134	488	99	523	5	13	120	648	92	704
3	8	38	386	23	428	6	12	174	736	141	791
4	7	69	474	54	495	7	11	233	815	192	869
5	6	130	526	100	568	8	10	296	891	250	943
4	8	73	469	60	519	9	9	364	962	314	1012
5	7	124	560	90	589	4	15	76	570	51	619
6	6	178	614	149	643	5	14	122	660	94	723
4	9	77	496	51	547	6	13	174	751	140	811
5	8	121	570	95	610	7	12	233	835	193	891
6	7	177	642	145	673	8	11	294	912	250	965
4	10	78	508	60	560	9	10	362	986	313	1039
5	9	123	593	98	645	2	18	9	392	3	449
6	8	175	664	142	702	7	13	35	481	22	556
7	7	241	727	199	769	4	16	73	578	53	639
4	11	74	523	52	573	5	15	120	673	92	737
5	10	123	605	93	652	6	4	172	763	139	823
6	9	177	683	144	729	7	13	230	849	190	907
7	8	238	753	199	794	8	12	293	930	247	986
4	12	73	532	53	581	11	9	359	1005	309	1061
5	11	122	627	97	667	10	10	429	1079	375	1133
6	10	177	720	140	751	8	8	301	845	257	889
7	9	235	776	197	820						

Если $|L^*| = \dfrac{|L - \mathbf{M}(L)|}{\mathbf{D}(L)} < u_{\frac{1+\alpha}{2}}$, нулевая гипотеза принимается.

Критерий Клотца оптимален для распределений, близких к нормальному.

Задача 265. *Проверить гипотезу равенства параметров масштаба критерием Клотца для данных задачи 262.*

Имеем последовательность рангов R_i: 2, 4, 7, 9, 10, 11, 13, 15.

Из табл. 162 находим для $N = m + n = 18$:

$$u^2_{\frac{2}{19}} = 262;\quad u^2_{\frac{4}{19}} = 65;\quad u^2_{\frac{7}{19}} = 11;\quad u^2_{\frac{9}{19}} = 0;$$
$$u^2_{\frac{10}{19}} = 0;\quad u^2_{\frac{11}{19}} = 4;\quad u^2_{\frac{13}{19}} = 23;\quad u^2_{\frac{15}{19}} = 65.$$

Вычисляем $L = 262 + 65 + 11 + 0 + 0 + 0 + 4 + 23 + 65 = 430$.

Из табл. 163 для $m = 8$, $n = 10$ и $\alpha = 0{,}95$ находим $L_1 = 250$ и $L_2 = 943$.

Так как $L_1 = 250 < L = 430 < L_2 = 943$, параметры масштаба двух выборок признаются статистически неразличимыми с достоверностью $\alpha = 0{,}95$.

Используем теперь нормальную аппроксимацию. Вычисляем, используя таблицу 162:

$$\mathbf{M}(L) = \frac{8}{18} \cdot (262 + 157 + 101 + 65 + \ldots + 157 + 262) = 589{,}333;$$

$$\mathbf{D}(L) = \frac{8 \cdot 10}{18 \cdot 17} \cdot \left(262^2 + 157^2 + 101^2 + 65^2 + \ldots + 157^+ 262^2\right) = 55280{,}087;$$

$$\sqrt{\mathbf{D}(L)} = 235{,}117; \quad L^* = \frac{|430 - 589{,}333|}{235{,}117} = 0{,}678,$$

что меньше $u_{0,975} = 1{,}96$; следовательно, и в этом случае нулевая гипотеза не отклоняется.

4.2.2.1.5. Квартильный критерий

Критерий является интуитивным аналогом медианного критерия сдвига (см. раздел 4.2.1.2.7). Статистика критерия имеет вид

$$S = \frac{1}{2}\sum_{i=1}^{m} \delta\left[\left(\left|R_i - \frac{m+n+1}{2}\right| - \frac{m+n+1}{2}\right)\right], \quad \text{где} \quad \delta(x) = \begin{cases} 1, & x > 0; \\ 0, & x = 0; \\ -1, & x < 0. \end{cases}$$

Название критерия исходит из того, что S приблизительно равно числу наблюдений из первой выборки, лежащих за пределами первой и третьей квартилей объединенной выборки. Точнее, S получается, если подсчитать количество наблюдений x_i $(i = 1, \ldots, m)$, для которых $R_i < \dfrac{m+n+1}{4}$ или $R_i > \dfrac{3\,(m+n+1)}{4}$, и, если $(m+n+1)$ делится на 4, прибавить $1/2$ в случае, когда $R_i = \dfrac{m+n+1}{4}$ или $R_i = \dfrac{3(m+n+1)}{4}$ для некоторого $i = 1, \ldots, m$, или прибавить 1 в случае, когда оба последних равенства имеют место для некоторых двух различных индексов i.

При $m, n > 20$ статистика S имеет приближенно нормальное распределение со средним $\mathbf{M}(S)$ и дисперсией $\mathbf{D}(S)$ [365], где

$$\mathbf{M}(S) = \begin{cases} \dfrac{m(m+n-1)}{2(m+n)} & \text{при } m+n = 4k-1; \\ \dfrac{2km}{m+n} & \text{при } m+n \neq 4k+1; \end{cases}$$

$$\mathbf{D}(S) = \begin{cases} \dfrac{mn[(m+n)^2 - 2(m+n) - 1]}{4(m+n)^2(m+n-1)} & \text{при } m+n = 4k-1; \\ \dfrac{2kmn(m+n-2k)}{(m+n)^2(m+n-1)} & \text{при } m+n \neq 4k-1. \end{cases}$$

Поэтому нулевая гипотеза равенства параметров масштаба принимается, если $|S^*| = \dfrac{|S - \mathbf{M}(S)|}{\sqrt{\mathbf{D}(S)}} < u_{\frac{1+\alpha}{2}}$, где α — доверительная вероятность.

Эффективность критерия по сравнению с F-критерием в случае нормального распределения невелика и равна $\approx 0{,}37$, поэтому им рекомендуется пользоваться при $m, n > 50$.

Задача 266. *Проверить гипотезу равенства параметров масштаба для данных задачи 262 квартильным критерием.*

Имеем ряд рангов R_i: 2, 4, 7, 9, 10, 11, 13, 15.

Вычисляем $\delta - \left(\left|R_1 - \dfrac{8+10+1}{2}\right| - \dfrac{8+10+1}{4}\right) = \delta(|2 - 9{,}5| - 4{,}75) = \delta(2{,}75) = 1.$

Далее по аналогии имеем

$$\delta(|R_2-9{,}5|-4{,}75)=\delta(0{,}75)=1; \qquad \delta(|R_3-9{,}5|-4{,}75)=\delta(-2{,}25)=-1;$$
$$\delta(|R_4-9{,}5|-4{,}75)=\delta(-4{,}25)=-1; \quad \delta(|R_5-9{,}5|-4{,}75)=\delta(-4{,}25)=-1;$$
$$\delta(|R_6-9{,}5|-4{,}75)=\delta(-3{,}25)=-1; \quad \delta(|R_7-9{,}5|-4{,}75)=\delta(-1{,}25)=-1;$$
$$\delta(|R_8-9{,}5|-4{,}75)=\delta(0{,}75)=1.$$

Тогда $S=\frac{1}{2}\cdot(1-1-1-1-1-1+1+8)=2{,}5$.

Далее вычисляем, имея в виду, что при $k=4m+n=4k+2 \quad (8+10=4\cdot 4+2)$

$$\mathbf{M}(S)=\frac{2\cdot 4\cdot 8}{8+10}=3{,}55; \quad \mathbf{D}(S)=\frac{2\cdot 4\cdot 8\cdot 10\cdot(8+10-8)}{18^2\cdot 17}=1{,}1619 \quad (\sqrt{\mathbf{D}(S)}=1{,}078).$$

Имеем $|S^*|=\dfrac{|2{,}5-3{,}55|}{1{,}078}=0{,}97$. Так как $|S^*|=0{,}97<u_{0,975}=1{,}96$, нулевая гипотеза принимается. Следует отметить, что к этому выводу нужно относиться с осторожностью. Это связано с тем, что точность аппроксимации при наших объемах выборок ($m=8$ и $n=10$) невелика.

4.2.2.1.6. Критерий Сэвиджа

Для распределений, плотность которых определена на полупрямой ($0\leqslant x<\infty$), т. е. экспоненциального типа, Сэвидж [484] предложил критерий масштаба, основанный на статистике

$$C=\sum_{i=1}^{m}\sum_{j=m+n-R_i+1}^{n}\frac{1}{j},$$

где R_i — так же, как и ранее, ранги элементов одной выборки объема $m\leqslant n$ в общем упорядоченном ряду.

Нулевая гипотеза равенства параметров масштаба принимается, если $C_1(\alpha)<C<C_2(\alpha)$, где $C_1(\alpha)$ и $C_2(\alpha)$ — критические значения, приведенные в табл. 164.

Применительно к экспоненциальному распределению критерий Сэвиджа позволяет установить разницу в значениях параметров распределений (λ или ν — см. раздел 1.1.4) в двух совокупностях (например, сравнить средние наработки на отказ изделий двух выборок).

Задача 267. *Даны две выборки случайных величин:*

$(m=7)$ x_{1j}: 2,1; 3,2; 4,8; 5,7; 8,1; 9,2; 12,4;
$(n=8)$ x_{2j}: 0,8; 0,9; 2,6; 4,3; 7,1; 9,4; 12,7; 15,1.

Проверить гипотезу равенства параметров масштаба в выборках критерием Сэвиджа при доверительной вероятности $\alpha=0{,}95$.

Ранжируем совместный ряд. Результаты сведем в таблицу (i — номер выборки, R_i — ранг члена выборки x_{ij} в объединенном ряду):

i	x_{ij}	R_i	i	x_{ij}	R_i	i	x_{ij}	R_i
2	0,8	1	2	4,3	6	1	9,2	11
2	0,9	2	1	4,8	7	2	9,4	12
1	2,1	3	1	5,7	8	1	12,4	13
2	2,6	4	2	7,1	9	2	12,7	14
1	3,2	5	1	8,1	10	2	15,1	15

Для первой выборки ($i=1$) получаем последовательность рангов

$$R_i=3,\ 5,\ 7,\ 8,\ 10,\ 11,\ 13.$$

Таблица 164

Критические значения $C_1(\alpha)$ и $C_2(\alpha)$ критерия Сэвиджа
(все значения умножены на 100, α — доверительная вероятность) [485]

m	n	α				m	n	α			
		0,90		0,95				0,90		0,95	
		C_1	C_2	C_1	C_2			C_1	C_2	C_1	C_2
3	3	116	485	116	485	2	14	42	431	33	493
2	5	45	418	45	418	3	13	99	571	79	619
3	4	121	494	96	527	4	12	165	685	140	736
2	6	40	444	40	444	5	11	240	798	207	849
3	5	119	507	83	566	6	10	321	905	283	955
4	4	205	595	171	629	7	9	407	1006	365	1054
2	7	49	416	35	466	8	8	498	1102	452	1148
3	6	110	521	90	566	2	15	45	438	31	505
4	5	190	623	165	654	3	14	99	572	79	622
2	8	44	436	31	486	4	13	164	691	137	743
3	7	109	534	92	571	5	12	237	805	205	859
4	6	188	636	157	677	6	11	317	913	279	967
5	5	2783	729	240	762	7	10	402	1017	359	1069
2	9	49	421	28	504	8	9	492	1116	445	1165
3	8	106	547	85	578	2	16	43	440	30	491
4	7	180	647	152	686	3	15	97	573	77	628
5	6	264	747	233	784	4	14	162	696	137	749
2	10	44	437	35	470	5	13	235	812	202	868
3	9	107	547	86	585	6	12	314	922	276	977
4	8	177	655	148	699	7	11	397	1027	355	1082
5	7	256	758	223	799	8	10	486	1128	439	1181
6	6	344	854	307	891	9	9	578	1225	528	1275
2	11	43	428	33	486	2	17	42	438	28	501
3	10	102	559	83	596	3	17	97	576	77	632
4	9	173	667	148	712	4	15	161	698	134	757
5	8	252	771	219	816	5	14	233	815	200	873
6	7	338	871	299	913	6	13	310	927	272	985
2	12	47	422	30	500	7	12	393	1034	349	1091
3	11	102	564	81	603	8	11	479	1136	432	1192
4	10	169	672	142	722	9	10	570	1235	520	1288
5	9	246	781	214	828	2	18	44	437	32	491
6	8	391	883	294	927	3	17	96	578	76	637
7	7	420	979	378	1021	4	16	160	702	133	761
2	13	44	435	36	480	5	15	230	820	198	880
3	12	101	563	81	613	6	14	307	933	269	993
4	11	167	677	143	727	7	13	389	1041	346	1101
5	10	243	791	212	839	8	12	475	1145	427	1204
6	9	325	895	288	943	9	11	564	1246	513	1302
7	8	414	994	372	1039	10	10	657	1342	603	1397

Вычисляем статистику критерия

$$C = \frac{1}{7+8-3+1} + \frac{1}{7+8-5+1} + \frac{1}{7+8-7+1} + \ldots + \frac{1}{7+8-13+1} = 1{,}104.$$

Из табл. 164 для $m = 7$, $n = 8$ и $\alpha = 0{,}95$ имеем (учитывая, что в ней приведены умноженные на 100 критические значения) $C_1 = 3{,}72$.

Так как $C = 1{,}104 < C_1 = 3{,}72$, нулевая гипотеза равенства параметров масштаба критерием Сэвиджа отклоняется.

Используем теперь нормальную аппроксимацию. При $m, n > 10$ справедливо нормальное приближение, в соответствии с которым величина распределена как стандартная нормальная случайная величина.

Для наших данных вычисляем

$$\mathbf{M}(C) = m = 7; \quad \mathbf{D}(C) = \frac{7 \cdot 8}{7+8-1} \cdot \left(1 - \frac{1}{7+8} \cdot \sum_{j=1}^{15} \frac{1}{j}\right) = 3{,}115 \quad (\sqrt{\mathbf{D}(C)} = 1{,}764).$$

Имеем $|C^*| = \dfrac{|1{,}104 - 7|}{1{,}764}$.

Так как $|C^*| = 3{,}342 > u_{0{,}975} = 1{,}96$, то и этот критерий отклоняет нулевую гипотезу.

4.2.2.1.7. Критерий Муда

Рассмотрен в [486, 487] в качестве альтернативы критерию, основанному на F-статистике Фишера, когда вместо наблюдений используются их ранги. Статистика критерия имеет вид

$$M = \sum_{i=1}^{m} \left(R_i - \frac{m+n+1}{2}\right)^2,$$

где R_i — ранги элементов выборки $x_1, \ldots, x_m$ в общем упорядоченном ряду значений $x_1, \ldots, x_m$ и $y_1, \ldots, y_n (m \leqslant n)$.

Нулевая гипотеза равенства параметров масштаба в обеих выборках принимается, если $m_1(\alpha) < M < m_2(\alpha)$, где $m_1(\alpha)$ и $m_2(\alpha)$ — критические значения статистики, приведенные в табл. 165.

При $m, n > 10$ справедлива нормальная аппроксимация [488, 489]

$$M^* = \frac{M - \mathbf{M}(M) + \dfrac{1}{2}}{\sqrt{\mathbf{D}(M)}},$$

где

$$\mathbf{M}(M) = \frac{m(m+n+1)(m+n-1)}{12}; \quad \mathbf{D}(M) = \frac{mn(m+n+1)(m+n+2)(m+n-2)}{180}.$$

Нулевая гипотеза принимается, если

$$|M^*| < u_{\frac{1+\alpha}{2}}.$$

Эффективность критерия Муда по отношению к F-критерию в случае исходного нормального распределения равна 0,76 [486].

Необходимо отметить, что критерий Муда (как и все ранее рассмотренные критерии) предполагает равенство средних (параметров положения). Это может быть обеспечено введением величин $x_i - \tilde{\mu}_1$ и $y_i - \tilde{\mu}_2$ вместо величин x_i и y_i ($\tilde{\mu}_1, \tilde{\mu}_2$ — медианы соответствующих выборок).

Задача 268. *Поверить гипотезу равенства параметров масштаба в условиях задачи 267 критерием Муда.*

Для ряда рангов R_i: 3, 5, 7, 8, 10, 11, 13 имеем

$$M = \sum_{i=1}^{m} \left(R_i - \frac{m+n+1}{2}\right)^2 = \left(3 - \frac{7+8+1}{2}\right)^2 + (5-8)^2 + (7-8)^2 + \ldots + (13-8)^2 = 73.$$

Из табл. 165 для $m = 7$, $n = 8$ и $\alpha = 0{,}95$ имеем $M_1 = 67$ и $M_2 = 196$.

Таблица 165

Критические значения $M_1(\alpha)$ и $M_2(\alpha)$ статистики Муда

(α — доверительная вероятность) [488]

m	n	α				m	n	α			
		0,90		0,95				0,90		0,95	
		M_1	M_2	M_1	M_2			M_1	M_2	M_1	M_2
2	6	1	18	1	18	6	10	65	191	55	203
2	7	1	25	1	25	6	11	74	217	62	230
2	8	2	32	1	32	6	12	81	245	69	259
2	9	2	34	1	41	6	13	90	274	75	291
2	10	2	50	1	50	6	14	99	303	83	323
2	11	3	57	1	61	7	7	68	160	60	168
2	12	3	72	2	72	7	8	76	185	67	196
2	13	4	74	2	80	7	9	86	214	74	224
3	3	3	15	3	15	7	10	95	243	83	255
3	4	5	22	2	22	7	11	106	274	92	290
3	5	5	27	3	31	7	12	116	307	100	324
3	6	5	34	5	36	7	13	128	342	110	362
3	7	9	43	5	47	8	8	106	234	94	246
3	8	9	51	5	57	8	9	117	267	104	280
3	9	9	63	7	71	8	10	130	302	116	318
3	10	11	77	8	81	8	11	144	338	127	356
3	11	13	87	9	93	8	12	158	378	138	398
4	4	9	33	9	33	9	9	156	328	140	344
4	5	11	42	9	45	9	10	172	369	155	386
4	6	15	53	11	55	5	10	42	147	34	155
4	7	15	66	14	70	5	11	47	107	39	179
4	8	19	79	15	83	5	12	54	190	43	203
4	9	21	93	15	101	6	6	41	101	35	107
4	10	23	107	19	117	6	7	46	123	39	130
5	5	21	61	17	65	6	8	51	143	43	151
5	6	25	76	20	79	6	9	59	166	50	176
5	7	29	91	23	95	9	11	188	410	168	432
5	8	33	107	27	114	10	10	220	444	200	464
5	9	37	125	31	135						

Так как $M_1 = 67 < M = 73 < M_2 = 196$, нулевая гипотеза не отклоняется.

Для нормального приближения имеем

$$\mathbf{M}(M) = \frac{7 \cdot 16 \cdot 14}{12} = 130{,}66; \quad \mathbf{D}(M) = \frac{7 \cdot 8 \cdot 16 \cdot 17 \cdot 13}{180} = 1100{,}08 \quad (\sqrt{\mathbf{D}(M)} = 33{,}16);$$

$$|M^*| = \frac{|73 - 130{,}61|}{33{,}16} = 1{,}737.$$

Так как $|M^*| = 1{,}737 < u_{0,975} = 1{,}96$, нулевая гипотеза этим критерием не отклоняется.

4.2.2.1.8. Критерий Сукхатме

Критерий является модифицированной формой критерия Вилкоксона–Уитни (см. раздел 4.2.1.1.2.2). Статистика критерия имеет вид [491, 492]

$$T = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \psi(x_i, y_j), \quad \text{где} \quad \psi(a, b) = \begin{cases} 1, & \text{если } 0 < a < b; \\ 0, & \text{если } a \leqslant 0, \quad a \geqslant b. \end{cases}$$

При $m, n > 10$ используется аппроксимация

$$|T^*| = \frac{|T - \mathbf{M}(T)|}{\sqrt{\mathbf{D}(T)}}, \quad \text{где} \quad \mathbf{M}(T) = \frac{1}{4}; \quad \mathbf{D}(T) = \frac{m+n+7}{48mn}.$$

При справедливости нулевой гипотезы T^* распределена как нормальная случайная величина. Поэтому с вероятностью α нулевая гипотеза отклоняется, если $|T^*| > u_{\frac{1+\alpha}{2}}$ (здесь u_γ — γ-квантиль стандартного нормального распределения).

Эффективность критерия Сукхатме для нормально распределенных случайных величин равна 0,61, для экспоненциально распределенных — 0,94.

Основным неудобством критерия, основанного на T-статистике, является необходимость предварительного знания соотношения параметров положения в выборках. Этого недостатка лишен модифицированный критерий, предложенный в [490] (имеем выборки $x_1, \ldots, x_m$ и $y_1, \ldots, y_n$):

$$T' = \sum_{i=1}^{n} \sum_{j \neq k}^{m} q(x_{j,k}, y_i) + 2 \sum_{i=p}^{n} \sum_{j=1}^{m} q(x_j, y_p, y_i) + \left(\frac{m+n}{2} - 1\right) \sum_{i=1}^{n} \sum_{j=1}^{m} K(x_j, y_i),$$

где

$$q(u, v, w) = \begin{cases} 1, & \text{если } 0 < u < w, \quad 0 < v < w \\ & \text{или } w < u < 0, \quad w < v < 0; \\ 0 & \text{в остальных случаях;} \end{cases}$$

$$K(u, v) = \begin{cases} 1, & \text{если } 0 < u < v, \\ & \text{или } v < u < 0; \\ 0 & \text{в остальных случаях.} \end{cases}$$

Более удобна для практических вычислений другая форма

$$T' = T_1' + T_2',$$

где

$$T_1' = \sum_{i=1}^{n_1} R_{i1}^2 + \frac{m+n-8}{2} \sum_{i=1}^{n_1} R_{i1} - \frac{n_1(n_1+1)[3(m+n) - 4n_1 - 22]}{12};$$

$$T_2' = \sum_{i=1}^{n_2} R_{i2}^2 - \frac{m+n+4(m_2+n_2)-4}{2} \sum_{i=1}^{n_2} R_{i2} + \frac{n_2(n_2+1)[3(m+n) + 8n_2 - 14]}{12}.$$

Здесь R_{i1} — ранг i-го положительного наблюдения y_i в упорядоченных по величине положительных значениях x_i и y_i; R_{i2} — ранг i-го отрицательного наблюдения в упорядоченном по величине ряду отрицательных значений x_i и y_i; n_1 — число положительных наблюдений y ($n_2 = n - n_1$); m_1 — число положительных наблюдений x ($m_2 = m - m_1$).

При справедливости нулевой гипотезы имеем

$$\mathbf{M}(T') = \frac{1}{24} n \left[5(m+n)^2 - 3n(m+n) - 2n^2 - 12(m+n) + 12n\right];$$

$$\mathbf{D}(T') = \frac{mn}{2880} \times$$
$$\times \left[61(m+n)^3 + 331(m+n)^2 - 120n^2 + 480n(m+n) - 2344(m+n) - 600n + 2636\right].$$

Эффективность T'-статистики по отношению к T-статистике равна 1,13.

Задача 269. *Проверить гипотезу масштаба критерием Сукхатме для данных задачи* 267.

Имеем выборки

$$x_i:\ 2{,}1,\ 3{,}2,\ 4{,}8,\ 5{,}7,\ 8{,}1,\ 8{,}2,\ 12{,}4 \qquad (m=7);$$
$$y_i:\ 0{,}8,\ 0{,}9,\ 2{,}6,\ 4{,}3,\ 7{,}1,\ 9{,}4,\ 12{,}7,\ 15{,}1 \qquad (n=8).$$

Значения $\psi(x_i, y_j)$ $(i = 1, \ldots, 7;\ j = 1, \ldots, 8)$ представлены в таблице:

i	j							
	1	2	3	4	5	6	7	8
1	0	0	1	1	1	1	1	1
2	0	0	0	1	1	1	1	1
3	0	0	0	0	1	1	1	1
4	0	0	0	0	1	1	1	1
5	0	0	0	0	0	1	1	1
6	0	0	0	0	0	1	1	1
7	0	0	0	0	0	0	1	1

Имеем

$$T = \frac{1}{m\cdot n}\cdot\sum_{i=1}^{m}\sum_{j=1}^{n}\psi(x_i, y_i) = \frac{27}{7\cdot 8} = 0{,}482; \quad \mathbf{D}(T) = \frac{7+8+7}{48\cdot 7\cdot 8} = 0{,}00818 \quad (\sqrt{\mathbf{D}(T)} = 0{,}09).$$

Окончательно имеем $|T^*| = \dfrac{|0{,}482 - 0{,}25|}{0{,}09} = 2{,}577$.

Так как $|T^*| = 2{,}577 > u_{0{,}975} = 1{,}96$, нулевая гипотеза отклоняется.

4.2.2.1.9. Критерий Сэндвика–Олссона

Критерий применим для парных выборок равного объема

$$x_1, x_2, \ldots, x_n;$$
$$y_1, y_2, \ldots, y_n.$$

Пусть m_x и m_y — оценки медиан в выборках. Построим последовательность величин

$$z_i = |x_i - m_x| - |y_i - m_y|, \quad i = 1, \ldots, n.$$

Нулевая гипотеза равенства параметров масштаба принимается, если z_i не имеют тенденции быть положительными [493]. Для проверки этого используется критерий знаковых рангов Вилкоксона (см. раздел 4.2.1.1.2.2).

Статистикой критерия является величина

$$T^+ = \sum_{z_i \geqslant 0} R_i,$$

где R_i — ранги неотрицательных значений z_i в общем ранжированном по возрастанию ряду $|z_i|$.

При $n > 20$ можно использовать приближение

$$T^* = \frac{T^+ - \dfrac{n(n+1)}{2}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}.$$

При $|T^*| > u_{\frac{1+\alpha}{2}}$ нулевая гипотеза отклоняется.

Задача 270. *Проверить критерием Сэндвика–Олссона равенство параметров масштаба для выборок*

x_i: 1,4, 1,8, 2,7, 3,9, 6,1, 7,8, 7,9;
y_i: 0,8, 0,9, 2,7, 5,9, 6,1, 8,1, 11,2

при доверительной вероятности $\alpha = 0{,}95$.

Для ряда x_i имеем медиану $m_x = 3{,}9$ и, по аналогии, $m_y = 5{,}9$. Вычисляем:

$$z_1 = |x_1 - m_x| - |y_1 - m_y| = |1{,}4 - 3{,}9| - |0{,}8 - 5{,}9| = -2{,}6;$$

$$z_2 = |1{,}8 - 3{,}9| - |0{,}9 - 5{,}9| = -2{,}9; \quad z_3 = |2{,}7 - 3{,}9| - |2{,}7 - 5{,}9| = -1{,}0;$$

$$z_4 = |3{,}9 - 3{,}9| - |5{,}9 - 5{,}9| = 0; \qquad z_5 = |6{,}1 - 3{,}9| - |6{,}1 - 5{,}9| = 2{,}0;$$

$$z_6 = |7{,}8 - 3{,}9| - (8{,}1 - 5{,}9) = 1{,}7; \qquad z_7 = |7{,}9 - 3{,}9| - |11{,}2 - 5{,}9| = -1{,}3.$$

Формируем упорядоченный ряд значений $|z_i|$ (внизу отмечены ранги):

$\lvert z_i \rvert$:	0	1,0	1,3	1,7	2,0	2,6	2,9
R_i:	1	2	3	4	5	6	7

Видим, что положительные значения $z_i(0; 1{,}7; 2{,}0)$ имеют ранги R_i: 1, 4 и 5 соответственно. Следовательно, $T^+ = 1 + 4 + 5 = 10$.

Из табл. 144 для $n = 7$ и $\alpha = 0{,}95$ имеем $T_1 = 4$ и $T_2 = 24$.

Так как $T_1 = 4 < T^+ = 10 < T_2 = 24$, нулевая гипотеза равенства параметров масштаба не отклоняется. Легко видеть, что к такому выводу приводит и аппроксимация

$$|T^*| = \frac{\left|10 - \dfrac{7 \cdot 8}{4}\right|}{\sqrt{\dfrac{7 \cdot 8 \cdot 15}{24}}} = 0{,}676 < u_{\frac{1+\alpha}{2}} = u_{0,975} = 1{,}96.$$

4.2.2.1.10. Критерий Краута–Линерта

Большинство из рассмотренных нами ранее критериев предполагали равенство параметров положения сравниваемых выборок. Краутом и Линертом [494, 495] рассмотрены критерии масштаба, нечувствительные к сдвигу. Если в нашем распоряжении находятся две зависимые (парные) выборки

$$x_i\colon\ x_1, x_2, \ldots, x_n;$$
$$y_i\colon\ y_1, y_2, \ldots, y_n,$$

то критерий масштаба строится следующим образом. Сначала случайным образом образуем $\frac{n}{2}$ пар совокупностей (x_{i_1}, y_{i_1}) и (x_{i_2}, y_{i_2}) и находим

$$D_i = \frac{(x_{i_1} - x_{i_2})^2 - (y_{i_1} - y_{i_2})^2}{2} \quad \left(1 \leqslant i \leqslant \frac{n}{2}\right).$$

Если совокупность величин D_i симметрична, то нулевая гипотеза равенства параметров масштаба не отклоняется. Таким образом, задача сводится к проверке симметричности распределения совокупности значений D_i. Это может быть выполнено одним из критериев, приведенных в главе 3 (см. раздел 3.5).

Для случая независимых выборок критерий строится следующим образом. Каждая выборка случайным образом разбивается на пары наблюдений

$$(x_{i_1}, x_{i_2}), \quad i = 1, \ldots, \frac{n}{2}, \qquad \text{и} \qquad (y_{j_1}, y_{j_2}), \quad j = 1, \ldots, \frac{n}{2}.$$

Затем производится сравнение рядов разностей

$$D_{x_i} = \frac{(x_{i_1} - x_{i_2})^2}{2} \quad \text{и} \quad D_{y_j} = \frac{(y_{j_1} - y_{j_2})^2}{2}$$

с помощью любого критерия сдвига.

Задача 271. *Для выборок равного объема $n = 16$*

x_i: 1, 2, 4, 6, 9, 12, 15, 21, 27, 31, 35, 46, 51, 61, 71, 75;

y_j: 2, 6, 9, 14, 21, 27, 38, 46, 59, 76, 91, 101, 110, 120, 140, 155

проверить гипотезу равенства параметров масштаба критерием Краута–Линерта при доверительной вероятности $\alpha = 0{,}95$.

Разбиваем выборку x_i произвольно на 8 пар:

$$(1{,}75),\quad (2{,}6),\quad (4{,}61),\quad (9{,}46),\quad (21{,}35),\quad (15{,}51),\quad (12{,}27),\quad (31{,}71).$$

Аналогично поступаем с выборкой y_j:

$$(6,9),\quad (2,27),\quad (21,155),\quad (14,101),\quad (38,59),\quad (46,140),\quad (91,120),\quad (76,110)$$

Вычисляем ряд значений

$$D_{x_i} = \frac{(x_{1i} - x_{2i})^2}{2} \to \frac{(1-75)^2}{2} = 2738;\quad \frac{(2-6)^2}{2} = 8;$$

$$\frac{(4-61)^2}{2} = 1624{,}5;\quad \frac{(9-46)^2}{2} = 684{,}5;\quad \frac{(21-35)^2}{2} = 98;$$

$$\frac{(15-51)^2}{2} = 648;\quad \frac{(12-27)^2}{2} = 112{,}5;\quad \frac{(31-71)^2}{2} = 800.$$

По аналогии получаем ряд

$$D_{y_i} = \frac{(y_{1i} - y_{2i})^2}{2} \to \frac{(6-9)^2}{2} = 4{,}5;\quad \frac{(2-27)^2}{2} = 312{,}5;$$

$$\frac{(21-155)^2}{2} = 8978;\quad \frac{(14-101)^2}{2} = 3784{,}5;\quad \frac{(38-59)^2}{2} = 220{,}5;$$

$$\frac{(46-140)^2}{2} = 4418;\quad \frac{(91-120)^2}{2} = 420{,}5;\quad \frac{(76-110)^2}{2} = 578.$$

Сдвиг рядов D_{x_i} и D_{y_i} проверяем критерием Вилкоксона (см. раздел 4.2.1.1.2.2), для чего строим ряд значений

$$z_1 = 2738 - 4{,}5 = 2733{,}5;\quad z_2 = 8 - 312{,}5 = -304{,}5;\quad z_3 = 1624{,}5 - 8978 = -7353{,}5;$$

$$z_4 = 684{,}5 - 3784{,}5 = -3100;\quad z_5 = 98 - 220{,}5 = -122{,}5;\quad z_6 = 648 - 4418 = -3770;$$

$$z_7 = 112{,}5 - 420{,}5 = -308;\quad z_8 = 800 - 578 = 222.$$

Ряд значений $|z_i|$ (внизу отмечены ранги) имеет вид:

$\lvert z_i\rvert \to$	122,5	222	304,5	308	2733,5	3100	3770	7353,5
$R_i \to$	1	2	3	4	5	6	7	8

Ранги R_i величин $z_i > 0$ (z_1 и z_8) в этом ряду: 2 и 5. Применяя критерий Сэндвика–Олсона (см. раздел 4.2.2.1.9), имеем

$$T^+ = \sum_{z_i > 0} R_i = 2 + 5 = 7.$$

Из табл. 144 для $n = 8$ и $\alpha = 0{,}95$ находим $T_1 = 4$ и $T_2 = 32$.

Так как $T_1 = 4 < T^+ = 7 < T_2 = 32$, гипотеза сдвига не подтверждается, а следовательно, не подтверждается различие параметров масштаба в выборках.

4.2.2.1.11. Критерий Камата

Является интуитивным масштабным аналогом критерия Хага (см. раздел 4.2.1.1.2.8). Его применение предполагает равенство параметров положения выборок (т. е. отсутствие сдвига между ними). Предварительно выборки $x_1, \ldots, x_n$ и $y_1, \ldots, y_m$ ($m \geqslant n$) совместно упорядочиваются. Элементы выборки x получают ранги R_{x_i}, а элементы выборки y — ранги R_{y_i}. Находим размахи рангов $R_n = (R_{x_i})_{\max} - (R_{y_i})_{\min}$ и $R_m = (R_{y_i})_{\max} - (R_{y_i})_{\min}$.

Статистика критерия Камата записывается как [496]

$$D_{n,m} = R_n - R_m + m.$$

Естественно, что $0 \leqslant D_{n,m} \leqslant m + n$. Нулевая гипотеза равенства параметров масштаба принимается, если $D_1(\alpha) < D_{n,m} < D_2(\alpha)$, где $D_1(\alpha)$ и $D_2(\alpha)$ — критические значения, приведенные в табл. 166.

Таблица 166

Критические значения $D_1(\alpha)$ и $D_2(\alpha)$ масштабного критерия Камата (α — доверительная вероятность) [496]

$m+n$	n	α				$m+n$	n	α			
		0,90		0,95				0,90		0,95	
		D_1	D_2	D_1	D_2			D_1	D_2	D_1	D_2
8	4	0	8			16	8	3	13	2	14
9	4	0	9			17	3	0	15	0	15
10	4	0	10	0	10		4	2	15	1	16
	5	0	10	0	10		5	2	15	2	16
11	3	0	11				6	3	15	2	15
	4	0	10	0	11		7	3	14	2	15
	5	1	10	0	11		8	3	14	3	15
12	3	0	11			18	3	1	16	0	16
	4	0	11	0	12		4	2	16	1	17
	5	1	11	0	12		5	3	16	2	16
	6	1	11	0	12		6	3	15	2	16
13	3	0	11				7	4	15	3	16
	4	1	12	0	13		8	4	15	3	15
	5	1	12	1	12		9	4	14	3	15
	6	1	12	1	12	19	3	1	16	0	17
14	3	0	12				4	2	16	1	18
	4	1	13	0	13		5	3	17	2	17
	5	2	13	1	13		6	4	16	3	17
	6	2	12	1	13		7	4	16	3	16
	7	2	12	1	13		8	4	15	3	16
15	3	0	13	0	15		9	4	15	3	16
	4	1	14	0	14	20	3	1	17	0	18
	5	2	13	1	14		4	2	17	1	19
	6	2	13	1	14		5	3	18	1	18
	7	2	13	2	13		6	4	17	3	18
16	3	0	14	0	15		7	5	17	4	17
	4	1	15	1	15		8	5	16	4	17
	5	2	14	1	15		9	4	16	4	16
	6	3	14	2	14		10	4	16	4	16
	7	3	13	2	14						

При $m + n > 20$ справедлива аппроксимация [496]

$$D_1(\alpha) = \mathbf{M}(D) - \sqrt{\mathbf{D}(D)}K_1(\alpha); \quad D_2(\alpha) = \mathbf{M}(D) + \sqrt{\mathbf{D}(D)}K_2(\alpha),$$

где

$$\mathbf{M}(D) = m - \frac{2m}{n+1} + \frac{2n}{m+1};$$

$$\mathbf{D}(D) = \frac{2m(n-1)(m+n+1)}{(n+1)^2(n+2)} + \frac{2n(m-1)(m+n+1)}{(m-1)^2(m+2)} + \frac{8mn}{(m+1)(n+1)} - 4 + 4(C_{m+n}^n)^{-1}.$$

Здесь $K_1(\alpha)$ и $K_2(\alpha)$ — коэффициенты, зависящие от и доверительной вероятности α (для $\alpha = 0{,}95$ их значения приведены в табл. 167).

Таблица 167

Значения корректирующих коэффициентов $K_1(0{,}95)$ и $K_2(0{,}95)$ критерия Камата

$P = \dfrac{m}{m+n}$	0,50	0,55	0,60	0,65	0,70	0,75
$K_2(0{,}95)$	1,99	1,88	1,77	1,67	1,57	1,50
$K_1(0{,}95)$	1,99	2,10	2,20	2,28	2,34	2,41

Задача 272. *Для данных задачи 270 проверить нулевую гипотезу равенства параметров масштаба критерием Камата.*

Запишем упорядоченный по возрастанию совместный ряд (i — номер выборки, R_i — ранг элемента в совместной последовательности значений x_i):

i	x_i	R_i	i	x_i	R_i	i	x_i	R_i
2	0,8	1	2	2,7	5,5	1	7,8	11
2	0,9	2	1	3,9	7	1	7,9	12
1	1,4	3	2	5,9	8	2	8,1	13
1	1,8	4	1	6,1	9,5	2	11,2	14
1	2,7	5,5	2	6,1	9,5			

Для первой выборки имеем последовательность рангов

$$R_{x_i} \to 3;\ 4;\ 5{,}5;\ 7;\ 9{,}5;\ 11;\ 12 \quad \text{и} \quad R_n = 12 - 3 = 9.$$

Для второй выборки имеем

$$R_{y_i} \to 1;\ 2;\ 5,5;\ 8;\ 9,5;\ 13;\ 14 \quad \text{и} \quad R_m = 14 - 1 = 13.$$

Тогда $D_{n,m} = 9 - 13 + 7 = 3$. Из табл. 166 имеем для $m + n = 14$, $m = 7$ и $\alpha = 0{,}95$: $D_1 = 1$ и $D_2 = 13$. Так как $D_1 = 1 < D_{n,m} = 3 < D_2 = 13$, нулевая гипотеза не отклоняется.

Рассмотрим теперь аппроксимацию. Имеем

$$\mathbf{M}(D) = 7 - \frac{2\cdot 7}{8} + \frac{2\cdot 7}{8} = 7;$$

$$\mathbf{D}(D) = \frac{2\cdot 7\cdot 6\cdot 15}{8^2\cdot 9} + \frac{2\cdot 7\cdot 6\cdot 15}{8^2\cdot 9} + \frac{8\cdot 7\cdot 7}{8\cdot 8} - 4 + \frac{4}{C_{14}^7} = 6{,}501 \quad (\sqrt{\mathbf{D}(D)} = 2{,}55).$$

Из табл. 167 для $p = 0{,}5$ имеем $K_1 = 1{,}99 = K_2$. Тогда $D_1 = 7 - 1{,}99\cdot 2{,}55 = 2$ и $D_2 = 7 + 1{,}99\cdot 2{,}55 = 12$.

Полученный диапазон критических значений несколько у́же точных (это следствие того, что $m + n < 20$), но и в этом случае нулевая гипотеза не отклоняется.

4.2.2.1.12. Комбинированный критерий Буша–Винда

Ранее мы рассматривали критерии сдвига и критерии масштаба, которые позволяли проверять гипотезы относительно одного из параметров (сдвига или масштаба), фиксируя условия для другого параметра (например, полагая их равенство). Однако на практике возможны случаи, когда нельзя априори знать, какие параметры сравниваемых выборок могут отличаться между собой. Естественно, в этом случае целесообразно иметь критерий, позволяющий одновременно проверить сдвиг

как в параметре положения, так и в масштабном параметре. Одним из таких ранговых комбинированных критериев является критерий Буша–Винда, рассмотренный в [497].

Статистика этого критерия строится следующим образом. Пусть мы имеем две независимые выборки $x_1, \ldots, x_m$ и $y_1, \ldots, y_n$ и R_i — ранги x_i в совокупной ранжированной по возрастанию выборке.

Введем метки

$$a_N(i) = u_{\frac{i}{N+1}} \quad \text{и} \quad b_N(i) = a_N^2(i)\,,$$

где $N = m + n$ и u_γ — γ-квантиль стандартного нормального распределения (можно использовать аппроксимацию для квантили $u_\gamma \approx 4{,}91\left[\gamma^{0,14} - (1-\gamma)^{0,14}\right]$, см. раздел. 1.1.1). Дисперсии меток a_N и b_N равны соответственно:

$$\mathbf{D}(a_N) = \frac{1}{N}\sum_{i=1}^{N} a_N^2(i); \quad \mathbf{D}(b_N) = \frac{1}{N}\sum_{i=1}^{N} b_N^2(i).$$

Введем статистики

$$S_{mn} = \sqrt{\frac{N-1}{mn}}\,\frac{\sum\limits_{i=1}^{m} a_N(R_i)}{\sqrt{\mathbf{D}(a_N)}}; \quad T_{mn} = \sqrt{\frac{N-1}{mn}}\,\frac{\sum\limits_{i=1}^{m} b_N(R_i) - \frac{m}{n}\sum\limits_{i=1}^{n} a_N^2(i)}{\sqrt{\mathbf{D}(b_N)}}.$$

Легко видеть, что статистика S_{mn} эквивалентна статистике Ван дер Вардена (см. раздел 4.2.1.1.2.4), а статистика T_{mn} — эквивалентна статистике Клотца (см. раздел 4.2.2.1.4). В качестве статистики комбинированного критерия предлагается статистика

$$W_{mn} = -2\ln[2\,(1 - \Phi\,(|S_{mn}|))] - 2\ln[2\,(1 - \Phi\,(|T_{mn}|))]\,,$$

где $\Phi(z)$ — функция стандартного нормального распределения.

Допустимо использование аппроксимации

$$\Phi(z) \approx 1 - 0{,}852\exp\left\{-\left(\frac{z + 1{,}5774}{2{,}0637}\right)^{2,34}\right\}.$$

Гипотеза отсутствия сдвига и неразличимости параметров масштаба с достоверностью α принимается, если $W_{mn} < W_{mn}(\alpha)$, где $W_{mn}(\alpha)$ — критические значения, приведенные в табл. 168.

Таблица 168

Критические значения статистики $W_{mn}(\alpha)$ Буша–Винда для $m = n$ [497]

$m = n$	Доверительная вероятность α			$m = n$	Доверительная вероятность α		
	0,90	0,95	0,975		0,90	0,95	0,975
5	7,03	7,92	8,57	20	7,65	9,05	10,50
6	7,19	8,06	8,67	25	7,69	9,09	10,50
7	7,31	8,20	9,06	30	7,71	9,24	10,60
8	7,36	8,47	9,25	40	7,74	9,30	10,70
9	7,45	8,54	9,50	50	7,76	9,32	10,90
10	7,48	8,65	9,65	∞	7,78	9,49	11,10
15	7,58	8,97	10,20				

При $m, n > 30$ справедлива аппроксимация $W_{mn}(\alpha) = \chi^2_\alpha(4)$, где $\chi^2_\alpha(4)$ — α-квантиль распределения хи-квадрат с $f = 4$ степенями свободы. Мощность критерия Буша–Винда весьма высока, что позволяет считать его эффективным критерием для непараметрической проверки гипотез в выборках из нормального (или близкого к нему) распределения.

Задача 273. *Для данных задачи 270 проверить гипотезу равенства параметров положения и масштаба в выборках комбинированным критерием Буша–Винда.*

Имеем ряд рангов R_i: 3; 4; 5,5; 7; 9,5; 11; 12.

Используя аппроксимацию

$$a_{14}(i) = u_{\frac{i}{14+1}} = 4{,}91 \cdot \left[\left(\frac{i}{15}\right)^{0,14} - \left(1 - \frac{i}{15}\right)^{0,14}\right],$$

вычисляем значения $a_{14}(i)$. Результаты вычислений сводим в таблицу (там же указаны величины $a^2_{14}(i)$ и дополнительные величины $a_{14}(i)$ для дробных рангов 5,5 и 9,5):

i	$a_{14}(i)$	$a^2_{14}(i)$	i	$a_{14}(i)$	$a^2_{14}(i)$
1	−1,5021	2,2563	8	0,0833	0,00693
2	−1,1094	1,2308	9	0,2522	0,0636
3	−0,8395	0,7048	9,5	0,3393	0,1151
4	−0,6208	0,3854	10	0,4290	0,1841
5	−0,4290	0,1841	11	0,6208	0,3854
5,5	−0,3393	0,1151	12	0,8395	0,7048
6	−0,2522	0,0636	13	1,1094	1,2308
7	−0,0833	0,00693	14	1,5021	2,2563

Вычисляем

$$\mathbf{D}(a_N) = \frac{1}{N} \cdot \sum_{i=1}^{N} a^2_N(i) = \frac{1}{14} \cdot \sum_{i=1}^{14} a^2_{14}(i) = 0{,}695 \quad (\sqrt{\mathbf{D}(a_N)} = 0{,}833);$$

$$s_{mn} = \sqrt{\frac{14-1}{7 \cdot 7}} \cdot \frac{\sum_{i=1}^{7} a_{14}(R_i)}{0{,}833} = 0{,}618 \cdot [a_{13}(3) + a_{14}(4) + a_{14}(5{,}5) + \ldots + a_{14}(12)] = -0{,}0515.$$

Далее вычисляем

$$\mathbf{D}(b_N) = \frac{1}{N} \cdot \sum_{i=1}^{N} b^2_N(i) = \frac{1}{N} \cdot \sum_{i=1}^{N} a^4_N(i) = 1{,}043 \quad (\sqrt{\mathbf{D}(b_N)} = 1{,}021);$$

$$T_{mn} = \sqrt{\frac{13}{49}} \cdot \frac{\sum_{i=1}^{7} a^2_N(R_i) - \frac{7}{14} \cdot \sum_{i=1}^{14} a^2_N(i)}{\sqrt{\mathbf{D}(b_N)}} = \frac{0{,}5151}{1{,}021} \cdot \left[\sum_{i=1}^{7} a^2_N(R_i) - 0{,}5 \cdot \sum_{i=1}^{14} a^2_N(i)\right] =$$

$$= 0{,}5048 \cdot \left[a^2_{14}(3) + a^2_{14}(4) + \ldots + a^2_{14}(12) - 0{,}5 \cdot a^1_{14}(1) - 0{,}5 \cdot a^2_{14}(2) - \ldots \right.$$
$$\left. \ldots - 0{,}5 \cdot a^2_{14}(14)\right] = -1{,}235.$$

Находим аппроксимации

$$\Phi(0{,}0515) = 1 - 0{,}852 \cdot \exp\left\{-\left(\frac{0{,}0515 - 1{,}5774}{2{,}0637}\right)^{2,34}\right\} = 0{,}5205;$$

$$\Phi(1{,}235) = 1 - 0{,}852 \cdot \exp\left\{-\left(\frac{1{,}235 + 1{,}5774}{2{,}0637}\right)^{2,34}\right\} = 0{,}8918.$$

Окончательно имеем

$$W_{mn} = -2 \cdot \ln[2 \cdot (1 - 0{,}5205)] - 2 \cdot \ln[2 \cdot (1 - 0{,}8918)] = 3{,}145.$$

Из табл. 168 для $\alpha = 0{,}95$ и $m = n = 7$ имеем $W_{mn}(0{,}95) = 8{,}20$.

Так как $W_{mn} = 3{,}145 < W_{mn}(0{,}95) = 8{,}20$, гипотеза равенства параметров положения и масштаба в выборках не отклоняется.

4.2.2.2. Сравнение параметров масштаба нескольких ($k > 2$) совокупностей критерием Бхапкара–Дешпанде

Рассмотрим совокупность k выборок

$$x_{11}, x_{12}, \ldots, x_{1n_1}; \quad x_{21}, x_{22}, \ldots, x_{2n_2}; \quad \ldots; \quad x_{i1}, x_{i2}, \ldots, x_{in_i}; \quad \ldots; \quad x_{k1}, x_{k2}, \ldots, x_{kn_k}.$$

Пусть x_{ij} — j-е наблюдение в i-й выборке ($j = 1, 2, \ldots, n_i$). Если мы будем всеми возможными способами из каждой выборки брать одно наблюдение, то можем получить $\prod\limits_{i=1}^{k} n_i$ подвыборок.

Пусть v_{ij} — число таких подвыборок, в которых наблюдение, соответствующее i-й выборке, было больше, чем $(j-1)$ наблюдений (или меньше, чем остальные $(k-j)$ наблюдений). Определим величины $u_{ij} = \dfrac{v_j^2}{\prod\limits_i n_i}$, которые распределены равномерно на интервале $[0, 1]$. Бхапкар [499] показал, что при справедливости нулевой гипотезы $H_0\colon F_1(x) = F(x) = \ldots = F_k(x)$, утверждающей, что распределения в сравниваемых выборках совпадают с точностью до параметров положения и масштаба, имеет место соотношение $u_{i1} = \dfrac{1}{k}$.

Там же Бхапкар предложил критерий для проверки сдвига как в дисперсиях, так и в средних сравниваемых выборок. В случае равенства параметров положения этот критерий является критерием проверки равенства параметров масштаба. Критерий Бхапкара основан на статистике

$$V = \prod_{i+1}^{k} n_i(2k-1)\left[\sum_{i=1}^{k}\frac{n_i}{\prod\limits_{i=1}^{k} n_i}\left(u_{i1}-\frac{1}{k}\right)^2 - \left\{\sum_{i=1}^{k}\frac{n_i}{\prod\limits_{i=1}^{k} n_i}\left(u_{i1}-\frac{1}{k}\right)\right\}^2\right].$$

Дешпанде, Дюфор и Люнг [500, 501] предложили эффективный критерий для выявления разницы в параметрах масштаба, основанный на статистике

$$D = \frac{\prod\limits_{i=1}^{k} n_i(k-1)^2(2k-1)C_{2(k-1)}^{k-1}}{2\left[k^2 + (k^2+4k+2)C_{2(k-1)}^{k-1}\right]}\left[\sum_{i=1}^{k}\frac{n_i}{\prod\limits_{i=1}^{k} n_i} d_i^2 - \left\{\sum_{i=1}^{k}\frac{n_i}{\prod\limits_{i=1}^{k} n_i} d\right\}^2\right],$$

где $d_i = u_{i1} + u_{ik}$ $(i = 1, 2, \ldots, k)$.

Ими же предложена статистика для выявления сдвига в k выборках (что полезно делать перед применением критериев масштаба). Эта статистика имеет вид

$$L = \frac{\prod\limits_{i=1}^{k} n_i(2k-1)(k-1)^2 C_{2(k-1)}^{k-1}}{2k^2\left(C_{2(k-1)}^{k-1} - 1\right)}\left[\sum_{i=1}^{k}\frac{n_i}{\prod\limits_{i=1}^{k} n_i} l_i^2 - \left\{\sum_{i=1}^{k}\frac{n_i}{\prod\limits_{i=1}^{k} n_i} l_i\right\}^2\right],$$

где $l_i = -u_{i1} + u_{ik}$ $(i = 1, 2, \ldots, k)$.

В [502] рассмотрена статистика, эквивалентная статистике Крускала–Уоллиса (см. раздел 4.2.1.2.1). Нулевая гипотеза не отклоняется критериями V, L, D, если $V, L, D < \chi^2_\alpha(k-1)$, где $\chi^2_\alpha(k-1)$ — α-квантиль распределения хи-квадрат с $f = k-1$ степенями свободы.

Для случая $k = 3$ критические значения критериев L и D (L_α и D_α) приведены в табл. 169.

Таблица 169

Критические значения L_α и D_α для $k = 3$ [498]

n_1	n_2	n_3	Доверительная вероятность α			
			0,99		0,95	
			L	D	L	D
2	2	2			5,33	10,00
2	2	3	6,48	11,43	5,39	8,00
2	2	4	7,33	13,33	6,10	5,80
2	3	3	7,40	10,00	5,85	5,59
2	3	4	7,14	9,95	5,80	4,54
2	4	4	7,47	8,30	5,55	5,48
3	3	3	7,21	9,38	6,22	4,80
3	4	4	7,82	6,75	6,34	4,18
4	4	4	8,20	7,15	6,15	4,60

Задача 274. *Предположим, что имеются $k = 3$ выборки по четыре наблюдения в каждой*

$$x_{1i}: \ 300, \ 400, \ 510, \ 600;$$
$$x_{2i}: \ 250, \ 440, \ 570, \ 900;$$
$$x_{3i}: \ 520, \ 610, \ 920, \ 1070.$$

Необходимо проверить гипотезу о равенстве параметров масштаба в выборках критериями V, L и D при доверительной вероятности $\alpha = 0{,}95$.

В нашем случае $n_i = 4$ и $k = 3$. Таким образом, мы можем получить $\prod_{i=1}^{3} n_i = 4^3 = 64$ подвыборки объема $k = 3$ каждая (номер столбца в подвыборках соответствует номеру выборки)

300, 250, 520	400, 250, 520	510, 250, 520	600, 250, 520	300, 440, 520
300, 250, 610	400, 250, 610	510, 250, 610	600, 250, 610	300, 440, 610
300, 250, 920	400, 250, 920	510, 250, 920	600, 250, 1070	300, 440, 920
300, 250, 1070	400, 250, 1070	510, 250, 1070	600, 250, 1070	300, 440, 1070

400, 440, 520	510, 440, 520	600, 440, 520	300, 570, 520	400, 570, 520
400, 440, 610	510, 440, 610	600, 440, 610	300, 570, 610	400, 570, 610
400, 440, 920	510, 440, 920	600, 440, 920	300, 570, 920	400, 570, 920
400, 440, 1070	510, 440, 1070	600, 440, 1070	300, 570, 1070	400, 570, 1070

510, 570, 520	600, 570, 520	300, 900, 520	400, 900, 520	510, 900, 520	600, 900, 520
510, 570, 610	600, 570, 610	300, 900, 610	400, 900, 610	510, 900, 610	600, 900, 610
510, 570, 920	600, 570, 920	300, 900, 920	400, 900, 920	510, 900, 920	600, 900, 920
510, 570, 1070	600, 570, 1070	300, 900, 1070	400, 900, 1070	510, 900, 1070	600, 900, 1070

Находим v_{i1} — число подвыборок, в которых наблюдения в i-й выборке были меньше, чем $k - j = 3 - 1 = 2$ наблюдения. Прямым перебором убеждаемся, что $v_{11} = 35$,

$v_{21} = 27$, $v_{31} = 2$. По аналогии находим v_{i3} — число подвыборок, в которых наблюдения из i-й выборки были больше, чем $3 - 1 = 2$ наблюдения из остальных выборок. Имеем $v_{13} = 3$, $v_{23} = 11$, $v_{33} = 50$.

Проверкой убеждаемся, что

$$v_{11} + v_{12} + v_{13} = 35 + 27 + 2 = 64 \quad \text{и} \quad v_{13} + v_{23} + v_{33} = 3 + 11 + 50 = 64.$$

Далее находим

$$u_{11} = \frac{v_{11}}{64} = \frac{35}{64} = 0{,}5469; \quad u_{21} = \frac{27}{64} = 0{,}4219; \quad u_{31} = \frac{2}{64} = 0{,}03125;$$
$$u_{13} = \frac{3}{64} = 0{,}0469; \quad u_{23} = \frac{11}{64} = 0{,}1719; \quad u_{33} = \frac{50}{64} = 0{,}78125;$$

$$d_1 = u_{11} + u_{13} = 0{,}5469 + 0{,}0469 = 0{,}5938; \quad d_2 = u_{21} + u_{23} = 0{,}4219 + 0{,}1719 = 0{,}5938;$$
$$d_3 = u_{31} + u_{33} = 0{,}03125 + 0{,}78125 = 0{,}8125;$$

$$l_1 = -u_{11} + u_3 = -0{,}5469 + 0{,}0469 = -0{,}5; \quad l_2 = -u_{21} + u_{23} = -0{,}4219 + 0{,}1719 = -0{,}25;$$
$$l_3 = -u_{31} + u_{33} = -0{,}03125 + 0{,}78125 = 0{,}75.$$

Вычисляем статистики критериев

$$V = 64 \cdot (2 \cdot 3 - 1) \cdot \left[\sum_{i=1}^{3} \left(u_{i1} - \frac{1}{3}\right)^2 \cdot \frac{4}{64} - \left\{\sum_{i=1}^{3} \frac{4}{64} \cdot \left(u_{i1} - \frac{1}{3}\right)\right\}^2\right] =$$
$$= 320 \cdot \left[\frac{\left(0{,}5469 - \frac{1}{3}\right)^2}{16} + \frac{\left(0{,}4219 - \frac{1}{3}\right)^2}{16} + \frac{\left(0{,}03125 - \frac{1}{3}\right)^2}{16} - \right.$$
$$\left. - \left\{\frac{1}{16} \cdot [(0{,}549 + 0{,}4219 + 0{,}03125 - 1)]\right\}^2\right] = 2{,}89;$$

$$L = \frac{64 \cdot 5 \cdot 4 \cdot C_4^2}{2 \cdot \left[9 + (9 + 12 + 2) \cdot C_4^2\right]} \cdot \left[\sum_{i=1}^{3} \frac{1}{16} \cdot d_i^2 - \left(\frac{1}{16} \cdot \sum_{i=1}^{3} d_i\right)^2\right] = 1{,}82.$$

Из табл. 169 для $\alpha = 0{,}95$ и $n_i = 4$ находим $L_{0,95} = 6{,}15$ и $D_{0,95} = 4{,}60$.

Так как $D = 1{,}82 < D_{0,95} = 4{,}60$ и $L = 4{,}67 < L_{0,95} = 6{,}15$, нулевая гипотеза не отклоняется.

К такому же результату приводит сравнение с $\chi^2_{0,95}(2)$, так как

$$V = 2{,}89, \quad L = 4{,}64, \quad D = 1{,}82 < \chi^2_{0,95}(2) = 5{,}99.$$

4.3. Критерии тренда и случайности

Критерии этого раздела предназначены для проверки гипотез о случайности расположения полученных выборочных данных, т. е. отсутствия взаимосвязи между значениями реализаций наблюдаемой случайной величины и их номерами в выборочной последовательности.

В приводимых ниже критериях используются выборочные значения случайной величины в порядке их появления (т. е. они образуют временной ряд).

Наибольшее применение критерии тренда находят при статистическом контроле и предупредительном регулировании технологических процессов в промышленности, позволяя заранее статистически обоснованно выявить намечающуюся тенденцию ухудшения качества продукции. Для медика наличие тренда в исследуемом ряду данных о заболеваниях является объективным критерием оценки надвигающейся эпидемии. Количество возможных ситуаций, в которых выявление тренда (закономерности, а не случайности появления ряда данных) дает практически полезную информацию, велико, и каждый инженер или исследователь повседневно встречается с необходимостью использовать критерии настоящего раздела в своей работе.

4.3.1. Критерий Аббе–Линника

Пусть $x_1, \ldots, x_n$ — ряд значений взаимно независимых нормально распределенных случайных величин с математическими ожиданиями $\mu_1, \ldots, \mu_n$ соответственно и одинаковыми (но неизвестными) дисперсиями. Проверяется гипотеза о том, что все выборочные значения принадлежат одной генеральной совокупности со средним μ: ($H_0\colon \mu_i = \mu$, $i = 1, 2, \ldots, n$) против альтернативы тренда

$$H_1\colon\ |\mu_{i+1} - \mu_i| > 0, \quad i = 1, 2, \ldots, n-1.$$

Статистика критерия Аббе–Линника имеет вид

$$q = \frac{1}{2}\frac{\sum\limits_{i=1}^{n-1}(x_{i+1} - x_i)^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}, \quad \text{где} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Если $q > q_\alpha$, то нулевая гипотеза случайности ряда $x_1, \ldots, x_n$ отклоняется с доверительной вероятностью α (критические значения q_α приведены в табл. 170).

При $n > 60$ справедлива аппроксимация, основанная на том, что случайная величина $Q^* = -(1-q)\sqrt{\dfrac{2n+1}{2-(1-q)^2}}$ имеет стандартное нормальное распределение. Поэтому нулевая гипотеза отклоняется, если $Q^* < u_{1-\alpha}$.

В заключение упомянем достаточно простой критерий, приписываемый Кохрану [9], с помощью которого можно установить изменение среднего статистического ряда скачком после первых n_i наблюдений ($n_i + n_{i+1} = n$, $i = 2, \ldots, n-1$). Его статистики имеют вид

$$\chi_i^2 = \frac{n_i(n-n_i)}{n}\frac{(\bar{x}_{i\text{л}} - \bar{x}_{i\text{п}})^2}{\bar{x}},$$

где $\bar{x}_{i\text{л}} = \dfrac{1}{n_i}\sum\limits_{i=1}^{n_i} x_i$; $\bar{x}_{i\text{п}} = \dfrac{1}{n-n_i}\sum\limits_{i=n_i+1}^{n} x_i$; $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$.

Таблица 170

Критические значения q_α критерия Аббе–Линника [25]

n	Доверительная вероятность α		n	Доверительная вероятность α		n	Доверительная вероятность α	
	0,95	0,99		0,95	0,99		0,95	0,99
4	0,3902	0,3128	23	0,6713	0,5479	42	0,7521	0,6655
5	0,4102	0,2690	24	0,6776	0,5562	43	0,7550	0,6659
6	0,4451	0,2808	25	0,6839	0,5639	44	0,7576	0,6622
7	0,4680	0,3070	26	0,6893	0,5713	45	0,7603	0,6659
8	0,4912	0,3314	27	0,6946	0,5784	46	0,7628	0,6693
9	0,5121	0,3544	28	0,6996	0,5850	47	0,7653	0,6727
10	0,5311	0,3759	29	0,7047	0,5915	48	0,7767	0,6757
11	0,5482	0,3957	30	0,7091	0,5975	49	0,7698	0,6787
12	0,5636	0,4140	31	0,7136	0,6034	50	0,7718	0,6814
13	0,5778	0,4309	32	0,7177	0,6089	51	0,7739	0,6842
14	0,5908	0,4466	33	0,7216	0,6141	52	0,7759	0,6869
15	0,6027	0,4611	34	0,7256	0,6193	53	0,7779	0,6896
16	0,6137	0,4746	35	0,7292	0,6242	54	0,7799	0,6924
17	0,6237	0,4872	36	0,7328	0,6290	55	0,7817	0,6949
18	0,6330	0,4989	37	0,7363	0,6337	56	0,7836	0,6974
19	0,5417	0,5100	38	0,7396	0,6381	57	0,7853	0,6999
20	0,6498	0,5203	39	0,7429	0,6425	58	0,7872	0,7024
21	0,6574	0,5301	40	0,7461	0,6467	59	0,7891	0,7049
22	0,6645	0,5393	41	0,7491	0,6508	60	0,7906	0,7071

Если $\chi_i^2 > \chi_\alpha^2(1)$, то изменение среднего скачком после первых n_i наблюдений признается значимым с достоверностью α, здесь $\chi_\alpha^2(1)$ — α-квантиль распределения хи-квадрат с $f = 1$ степенью свободы.

Задача 275. *Имеется выборочный ряд значений случайных величин* ($n = 10$)

$$x_i\colon\ 4{,}3;\ 2{,}1;\ 0{,}9;\ 5{,}2;\ 4{,}8;\ 1{,}2;\ 0{,}8;\ 3{,}0;\ 6{,}1;\ 10{,}2.$$

Проверить гипотезу случайности ряда x_i критерием Аббе–Линника при доверительной вероятности $\alpha = 0{,}95$.

Вычисляем

$$\bar{x} = \frac{1}{10} \cdot \sum_{i=1}^{10} x_i = 3{,}86;$$

$$\sum_{i=1}^{9} (x_{i+1} - x_i)^2 = (2{,}1 - 4{,}3)^2 + (0{,}9 - 2{,}1)^2 + \ldots + (10{,}2 - 6{,}1)^2 = 69{,}31;$$

$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 77{,}124; \quad q = \frac{1}{2} \cdot \frac{69{,}31}{77{,}124} = 0{,}4493.$$

Из табл. 170 при $n = 10$ и $\alpha = 0{,}95$ имеем $q_{0,95} = 0{,}5311$.

Так как $q = 0{,}4493 < q_{0,95} = 0{,}5311$, нулевая гипотеза отклоняется, и ряд значений x_i с достоверностью 0,95 может быть признан случайным.

Для нормальной аппроксимации имеем

$$Q^* = -(1 - 0{,}4493)\sqrt{\frac{21}{2 - (1 - 0{,}4493)^2}} = -1{,}938.$$

Так как $Q^* = -1{,}938 < u_{0,05} = -1{,}645$, то этим критерием нулевая гипотеза отклоняется.

Теперь применим критерий Кохрана. Для последовательности пар значений n_i и n_{i+1}, равных $(2,8)$, $(3,7)$, $(4,5)$, $(5,5)$, $(6,4)$, $(7,3)$, $(8,2)$, находим значения:

$$\bar{x}_{2л} = \frac{4{,}3+2{,}1}{2} = 3{,}2 \quad \text{и} \quad \bar{x}_{2п} = \frac{0{,}9+5{,}2+4{,}8+1{,}2+0{,}8+3{,}0+6{,}1+10{,}2}{8} = 4{,}025;$$

$$\bar{x}_{3л} = \frac{4{,}3+2{,}1+0{,}9}{3} = 2{,}433 \quad \text{и} \quad \bar{x}_{3п} = \frac{5{,}2+4{,}8+1{,}2+0{,}8+3{,}0+6{,}1+10{,}2}{7} = 4{,}471;$$

$$\bar{x}_{4л} = 3{,}125 \quad \text{и} \quad \bar{x}_{4п} = 4{,}35; \quad \bar{x}_{5л} = 3{,}46 \quad \text{и} \quad \bar{x}_{5п} = 4{,}26; \quad \bar{x}_{6л} = 3{,}083 \quad \text{и} \quad \bar{x}_{6п} = 5{,}025;$$

$$\bar{x}_{7л} = 2{,}757 \quad \text{и} \quad \bar{x}_{7п} = 6{,}43; \quad \bar{x}_{8л} = 2{,}787 \quad \text{и} \quad \bar{x}_{8п} = 8{,}15.$$

Далее вычисляем

$$\chi_2^2 = \frac{2\cdot 8}{10}\cdot\frac{(\bar{x}_{2л}-\bar{x}_{2п})^2}{\bar{x}} = 1{,}6\cdot\frac{(3{,}2-4{,}025)^2}{3{,}86} - 0{,}282;$$

$$\chi_3^2 = 2{,}264; \quad \chi_4^2 = 0{,}933; \quad \chi_5^2 = 0{,}414; \quad \chi_6^2 = 2{,}344; \quad \chi_7^2 = 7{,}339; \quad \chi_8^2 = 11{,}92.$$

Из табл. 55 находим, что $\chi^2_{0,95}(1) = 3{,}481$ и, следовательно, критерий Кохрана признает скачок среднего после $n_i = 7$ наблюдений.

4.3.2. Критерий Фостера–Стюарта

Предложен в [503] и используется для проверки тренда как средних, так и дисперсий. Статистики критерия имеют вид

$$S = \sum_{i=2}^{n} S_i; \quad d = \sum_{i=2}^{n} d_i,$$

где

$$d_i = u_i - l_i; \quad S_i = u_i + l_i;$$

$$u_i = \begin{cases} 1, & \text{если } x_i > x_{i-1}, x_{i-2}, \ldots, x_1; \\ 0, & \text{в противном случае;} \end{cases} \qquad l_i = \begin{cases} 1, & \text{если } x_i < x_{i-1}, x_{i-2}, \ldots, x_l; \\ 0, & \text{в противном случае.} \end{cases}$$

Статистика S используется для проверки тренда в дисперсиях, статистика d — для обнаружения тренда в средних. Очевидно, что

$$0 \leqslant S \leqslant n-1 \quad \text{и} \quad -(n-1) \leqslant d \leqslant n-1.$$

При отсутствии тренда величины

$$t = \frac{d}{f} \quad \text{и} \quad \tilde{t} = \frac{S-f^2}{l}, \quad \text{где} \quad l = \sqrt{2\ln n - 3{,}4253}, \quad f = \sqrt{2\ln n - 0{,}8456},$$

имеют распределение Стьюдента с $\nu = n$ степенями свободы. Формулы для f и l применимы при $n > 50$, их значения при $n < 50$ приведены в табл. 171.

Таблица 171

Постоянные f и l критерия Фостера–Стюарта [503]

n	10	15	20	25	30	35	40	45	50
f	1,964	2,153	2,279	2,373	2,447	2,509	2,561	2,606	2,645
l	1,288	1,521	1,677	1,791	1,882	1,956	2,019	2,072	2,121

Если $|t|$ $(|\tilde{t}|) > t_{\frac{1+\alpha}{2}}$, то с доверительной вероятностью α нулевая гипотеза отсутствия трендов отклоняется (t_γ — γ-квантиль распределения Стьюдента).

Задача 276. *Проверить гипотезу наличия тренда критерием Фостера–Стюарта для данных задачи* 275.

Анализируя ряд x_i, получаем значения переменных u_i, l_i, d_i и S_i.
Результаты сводим в таблицу:

i	u_i	l_i	d_i	S_i	i	u_i	l_i	d_i	S_i
2	0	1	−1	1	7	0	1	−1	1
3	0	1	−1	1	8	0	0	0	0
4	1	0	1	1	9	1	0	1	1
5	0	0	0	0	10	1	0	1	1
6	0	0	0	0					

Далее находим $S = \sum\limits_{i=2}^{10} S_i = 6$ и $d = \sum\limits_{i=1}^{10} d_i = 0$. Из табл. 171 для $n = 10$ находим $f = 1{,}964$ и $l = 1{,}288$.

Далее имеем $t = \dfrac{0}{1{,}964} = 0$ и $\tilde{t} = \dfrac{6 - 1{,}964^2}{1{,}288} = 1{,}663$.

Для $t_{\frac{1+0{,}95}{2}} = t_{0{,}975}(\nu = 10) = 2{,}228$ (используем таблицы t-распределения [24–26]) получаем $|t| = 0$ ($|\tilde{t}| = 1{,}663$) $< t_{0{,}975} = 2{,}228$. Следовательно, наличие тренда не подтверждается имеющимися наблюдениями.

4.3.3. Критерий Кокс–Стюарта

В [504] предложена серия быстрых знаковых критериев тренда среднего и дисперсии в последовательности наблюдений. Для критерия среднего в выборке объема n предложена статистика

$$S_1 = \sum_{i=1}^{\left[\frac{n}{2}\right]} (n - 2i + 1)\, h_{i,n-i+1}, \quad \text{где} \quad h_{i,j} = \begin{cases} 1, & \text{если } x_i > x_j, \\ 0, & \text{если } x_i \leqslant x_j \end{cases} \quad (i < j).$$

Критерий, основанный на статистике S_1, имеет эффективность $\approx 0{,}86$ по отношению к наилучшему параметрическому критерию.

Для проверки гипотезы тренда применяется нормализованная статистика $S_1^* = \dfrac{S_1 - \mathbf{M}(S_1)}{\sqrt{\mathbf{D}(S_1)}}$, где $\mathbf{M}(S_1) = \dfrac{n^2}{8}$ и $\mathbf{D}(S_1) = \dfrac{n(n^2-1)}{24}$.

При $|S_1^*| < u_{\frac{1+\alpha}{2}}$ гипотеза тренда среднего отклоняется.

Критерий для проверки гипотезы о тренде дисперсии в выборке строится следующим образом. Выборка $x_1, \ldots, x_n$ разбивается на n/k подвыборок $x_1, \ldots, x_k$; $x_{k+1}, \ldots, x_{2k}$; $x_{2k+1}, \ldots, x_{3k}$; $x_{n-k+1}, \ldots, x_n$ (если n не делится на k, отбрасывается необходимое число наблюдений в центре). Для каждой i-й подвыборки находится размах ω_i $(1 \leqslant i \leqslant r)$ $\left(r = \left[\dfrac{n}{k}\right]\right)$. Далее размахи ω_i проверяются на тренд критерием S_1. Рекомендуется выбирать k из следующих соотношений [504]:

$$n \geqslant 90 \to k = 5; \quad 90 > n \geqslant 64 \to k = 4; \quad 64 > n \geqslant 48 \to k = 3; \quad n < 48 \to k = 2.$$

Эффективность дисперсионного критерия $\approx 0{,}73$.

Задача 277. *Имеется выборочная последовательность* ($n = 48$):

i:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_i:	12	1	2	4	6	9	5	3	14	21	24	29	1	3	7	2

i:	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
x_i:	13	21	14	51	1	0	0	17	21	3	4	6	7	8	9	1

i:	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
x_i:	2	3	12	11	10	21	74	35	28	24	15	0	1	17	19	1.

Необходимо проверить критериями Кокс–Стюарта гипотезу о тренде среднего и дисперсии в выборке при доверительной вероятности $\alpha = 0{,}95$.

Имеем значения $h_{i,n-i+1}$ для разных i ($i = 1, \ldots, 24$):

i	$n-i+1$	$h_{i,n-i+1}$	$n-i+1$	$h_{i,n-i+1}$	$n-i+1$	$h_{i,n-i+1}$	$n-i+1$	$h_{i,n-i+1}$
1	48	1	9	40	0	17	32	1
2	47	0	10	39	0	18	31	1
3	46	0	11	38	1	19	30	1
4	45	1	12	37	1	20	29	1
5	44	1	13	36	0	21	28	0
6	43	0	14	35	0	22	27	0
7	42	0	15	34	1	23	26	0
8	41	0	16	33	0	24	25	0

Находим

$$S_1 = \sum_{i=1}^{24} (49 - 2i) \cdot h_{i,n-i+1} = 47 + 41 + 39 + 27 + 25 + 19 + 15 + 13 + 11 + 9 = 246.$$

$$\mathbf{M}(S_1) = \frac{48^2}{8} = 288; \quad \mathbf{D}(S_1) = \frac{48 \cdot (48^2 - 1)}{24} = 4606 \quad (\sqrt{\mathbf{D}\,(S_1)} = 67{,}867);$$

$$|S_1^*| = \frac{|246 - 288|}{67{,}867} = 0{,}619.$$

Так как $|S_1^*| = 0{,}619 < u_{0,975} = 1{,}96$, гипотеза тренда отклоняется.

Рассмотрим дисперсионный критерий. Выбираем $k = 3$ и получаем $n/k = 16$ подвыборок:

12, 1, 2; 4, 6, 9; 5, 3, 14; 21, 24, 29; 1, 3, 7; 2, 13, 21; 14, 51, 1; 0, 0, 17;
21, 3, 4; 6, 7, 8; 9, 1, 2; 3, 12, 11; 10, 21, 74; 35, 28, 24 ; 15, 0, 1 ; 17, 19, 1.

Размахи подвыборок образуют ряд

10, 5, 11, 8, 6, 19, 50, 17, 18, 2, 8, 9, 64, 11, 15, 18.

Для полученного ряда находим значения $h_{i,17-i}$:

$$h_{1,16} = 0, \quad h_{2,15} = 0, \quad h_{3,14} = 0, \quad h_{4,13} = 0,$$
$$h_{5,12} = 0, \quad h_{6,11} = 1, \quad h_{7,10} = 1, \quad h_{8,9} = 0.$$

Тогда

$$S_1 = \sum_{i=1}^{8} (17 - i) \cdot h_{i,17-i} = 8; \quad \mathbf{M}(S_1) = \frac{16^2}{8} = 32; \quad \mathbf{D}(S_1) = \frac{16 \cdot (16^2 - 1)}{24} = 170;$$

$$\sqrt{\mathbf{D}(S_1)} = 130{,}38; \quad |S_1^*| = \frac{|8 - 32|}{13{,}038} = 1{,}84.$$

Так как $|S_1^*| = 1{,}84 < u_{0,975} = 1{,}96$, гипотеза тренда отклоняется.

4.3.4. Критерий обнаружения сдвига дисперсии в неизвестной точке (критерий Хсу)

Предположим, что наблюдаемая нами последовательность $x_1, \ldots, x_n$ имеет одно и то же среднее (оцениваемое медианой ряда m_x). Необходимо проверить нулевую гипотезу H_0 о неизменности дисперсии в выборке против альтернативы H_1, утверждающей, что значение дисперсии меняется в неизвестной точке.

Условно это можно записать так: проверить гипотезу $H_0\colon \sigma_1^2 = \ldots = \sigma_n^2 = \sigma_0^2$ (σ_0^2 неизвестно) против гипотезы

$$H_1\colon \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma_0^2; \quad \sigma_{k+1}^2 = \sigma_{k+2}^2 = \ldots = \sigma_n^2 = \sigma_0^2 + \delta \ \ (|\delta| > 0)\,,$$

где k неизвестно ($1 \leqslant k \leqslant n-1$).

В [505] предложен критерий для проверки такой гипотезы, основанный на статистике

$$H = \frac{\sum\limits_{i=1}^{n} (i-1)(x_i - m_x)^2}{(n-1)\sum\limits_{i=1}^{n} (x_i - m_x)^2}, \quad 0 \leqslant H \leqslant 1.$$

Обычно критерий используется в стандартизированной форме

$$H^* = \frac{H - 0{,}5}{\sqrt{\mathbf{D}(H)}}, \quad \text{где} \quad \mathbf{D}(H) = \frac{n+1}{6(n-1)(n+2)}.$$

Если $H^* < H_\alpha$, то с достоверностью α гипотеза изменения дисперсии отклоняется (H_α — критические значения, приведенные в табл. 172). При $n > 30$ применима нормальная аппроксимация $H_\alpha \approx U_\alpha$.

Таблица 172

Критические значения (H_α) H^*-критерия Хсу [505]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
2	1,345	1,397	1,414	9	1,301	1,650	2,266
3	1,370	1,613	1,901	10	1,299	1,650	2,272
4	1,343	1,643	2,083	15	1,294	1,648	2,289
5	1,325	1,650	2,170	20	1,290	1,647	2,298
6	1,314	1,652	2,219	25	1,289	1,647	2,304
7	1,306	1,652	2,251	30	1,287	1,646	2,308
8	1,303	1,651	2,260	∞	1,282	1,645	2,326

Определить точку изменения дисперсии позволяет G-критерий, также рассмотренный в [505]. Его статистика строится следующим образом.

Пусть

$$\omega_k = \sum_{i=1}^{k} (x_i - m_x)^2 \quad \text{и} \quad W_k = \frac{\omega_n - \omega_k}{\omega_k}\frac{k}{n-k}, \quad k = 1, 2, \ldots, n-1,$$

где k соответствует искомой точке изменения дисперсии.

Пусть далее α_k есть вероятность, соответствующая условию $F_{\alpha_k}(n-k, k) = W_k$, где $F_\gamma(f_1, f_2)$ — γ-квантиль F-распределения с f_1 и f_2 степенями свободы. Тогда

статистика искомого критерия есть

$$G = \sum_{k=1}^{n-1} \frac{\alpha_k}{n-1}, \quad 0 \leqslant G \leqslant 1.$$

Гипотеза изменения дисперсии отклоняется с вероятностью α, если $G < G_\alpha$, где G_α — критические значения, приведенные в табл. 173.

Таблица 173

Критические значения (G_α) G-критерия Хсу [505]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
2	0,895	0,917	0,989	9	0,781	0,840	0,921
3	0,838	0,898	0,964	10	0,779	0,838	0,920
4	0,813	0,874	0,948	15	0,774	0,833	0,915
5	0,800	0,860	0,938	20	0,772	0,831	0,913
6	0,792	0,852	0,931	30	0,770	0,829	0,911
7	0,787	0,847	0,927	100	0,768	0,826	0,909
8	0,783	0,843	0,926	∞	0,767	0,825	0,908

Значение k, для которого величина $\left|\alpha_k - \frac{1}{2}\right|$ максимальна, дает оценку точке изменения значения дисперсии в наблюдаемом ряду. Мощность критериев H и G близка.

Задача 278. *Для ряда значений*

$$x_i\colon\ 2,\ 1,\ 11,\ 21,\ 3,\ 8,\ 6,\ 23,\ 28,\ 38,\ 37\ (n = 10)$$

проверить гипотезу о возможном изменения дисперсии ряда критериями Хсу при доверительной вероятности $\alpha = 0{,}95$.

Находим медиану ряда (напомним, что это средний член ранжированного по возрастанию ряда, если n — нечетное, и полусумма центральных значений, если n — четное). В нашем случае $m_x = \dfrac{8+11}{2} = 9{,}5$.

Далее вычисляем

$$H = \frac{\sum\limits_{i=1}^{10} (i-1)\cdot(x_i - 9{,}5)^2}{(10-1)\cdot\sum\limits_{i=1}^{10}(x_i - 9{,}5)^2} = \frac{1\cdot(1-9{,}5)^2 + 2\cdot(11-9{,}5)^2 + \ldots + 9\cdot(37-9{,}5)^2}{9\cdot[(2-9{,}5)^2 + (1-9{,}5)^2 + \ldots + (37-9{,}5)^2]} = 0{,}831;$$

$$\mathbf{D}(H) = \frac{11}{6\cdot 9\cdot 12} = 0{,}01697 \quad (\sqrt{\mathbf{D}(H)} = 0{,}130); \quad H^* = \frac{|0{,}831 - 0{,}5|}{0{,}130} = 2{,}546.$$

Из табл. 172 для $\alpha = 0{,}95$ и $n = 10$ находим $H_{0,95} = 1{,}65$.

Так как $H^* = 2{,}546 > H_{0,95} = 1{,}65$, нулевая гипотеза отклоняется в пользу альтернативы изменения дисперсии ряда.

Вычислим теперь статистику G-критерия. Находим значения ω_i и W_i (результаты сведем в таблицу).

Вероятности α_k находим из соотношений

$$F\alpha_1(9{,}1) = 3{,}979; \quad F\alpha_2(8{,}2) = 3{,}778; \quad F\alpha_3\,(7{,}3) = 6{,}358;$$
$$F\alpha_4(6{,}4) = 4{,}582; \quad F\alpha_5(5{,}5) = 5{,}783; \quad F\alpha_6(4{,}6) = 8{,}60;$$
$$F\alpha_7(3{,}7) = 12{,}776; \quad F\alpha_8\,(2{,}8) = 12{,}498; \quad F\alpha_9(1{,}9) = 5{,}179.$$

i	ω_i	W_i	i	ω_i	W_i	i	ω_i	W_i
1	56,25	3,979	5	302,25	5,783	9	1314,25	5,179
2	128,50	3,778	6	307,50	8,600	10	2070,50	
3	130,75	6,358	7	319,75	12,776			
4	263,00	4,582	8	502,00	12,498			

Следует отметить,что имеющиеся таблицы F-распределения не столь обширны, чтобы по ним можно было быстро определить значения α_k. Мы это сделаем интерполяцией F_α, однако именно из-за отсутствия и громоздкости таких таблиц G-критерий вряд ли может быть рекомендован к широкому применению. Тем не менее, используя наиболее широкие таблицы F-распределения и применяя интерполяцию, находим:

$$\alpha_1 = 0{,}5675; \quad \alpha_2 = 0{,}7606; \quad \alpha_3 = 0{,}9151; \quad \alpha_4 = 0{,}9137; \quad \alpha_5 = 0{,}9603; \quad \alpha_6 = 0{,}9872;$$

$$\alpha_7 = 0{,}9595; \quad \alpha_8 = 0{,}9995; \quad \alpha_9 = 0{,}9507 \quad \text{и} \quad G = \frac{1}{n-1}\cdot\sum_{i=1}^{9}\alpha_i = 0{,}894.$$

Так как $G = 0{,}894 > G_{0,95} = 0{,}838$ (см. табл. 173), этот критерий отклоняет нулевую гипотезу.

Рассмотрим теперь ряд

$$\left|\alpha_k - \frac{1}{2}\right| \to 0{,}0675;\ 0{,}2606;\ 0{,}4151;\ 0{,}4137;\ 0{,}4603;\ 0{,}4872;\ 0{,}4595;\ 0{,}4995;\ 0{,}4507.$$

Из полученного ряда следует, что k: $\left(\left|\alpha_k - \frac{1}{2}\right| = \max\right) = 8$.

Следовательно, начиная с 8-го члена ряда, дисперсия начинает меняться.

4.3.5. Ранговый критерий обнаружения сдвига дисперсии в неизвестной точке

Критерий Хсу (см. раздел 4.3.4) предполагает нормальность распределения случайной величины. В [506] рассмотрен критерий для решения аналогичной задачи, но свободный от распределения. Он основан на использовании семейства ранговых статистик вида

$$S = \sum_{i=1}^{n} i a_n(R_i),$$

где R_i — ранги выборочных значений в упорядоченном ряду наблюдений.

Метки критерия a_n могут быть различными, например

— метки Клотца (см. раздел 4.2.2.1.4) $a_{1n}(i) = U^2_{\frac{i}{n+1}}$;

— метки Сэвиджа (см. раздел 4.2.2.1.6) $a_{2n}(i) = \sum_{j=1}^{i}\frac{1}{n-j+1}$.

Обозначим $S_j = \sum_{i=1}^{n} i a_{jn}(R_i)$, $j = 1, 2$. При отсутствии сдвига дисперсии в ряду наблюдений S_j-статистики свободны от распределения и симметричны относительно $\mathbf{M}(S_j) = \frac{n+1}{2}\sum_{i=1}^{n} a_{jn}(i)$. При справедливости нулевой гипотезы

$$\mathbf{M}(S_1) = \frac{n+1}{2}\sum_{i=1}^{n} U^2_{\frac{i}{n+1}}; \quad \mathbf{M}(S_2) = \frac{n(n+1)}{2};$$

$$\mathbf{D}(S_1) = \frac{n(n+1)}{12}\sum_{i=1}^{n} U^4_{\frac{i}{n+1}} - \frac{1}{3n+3}[M(S_1)]^2; \quad \mathbf{D}(S_2) = \frac{n(n+1)}{12}\left(n - \sum_{j=1}^{i}\frac{1}{j}\right).$$

Статистики критериев имеют вид

$$S_j^* = \frac{S_j - \mathbf{M}(S_j)}{\sqrt{\mathbf{D}(S_j)}}.$$

Нулевая гипотеза отклоняется с доверительной вероятностью α, если $|S_j^*| < S_j(\alpha)$, где $S_j(\alpha)$ — критические значения, приведенные в табл. 174.

Таблица 174

Критические значения рангового критерия обнаружения сдвига дисперсии [506]

n	Доверительная вероятность $\alpha = 0{,}95$		Доверительная вероятность $\alpha = 0{,}90$	
	S_1	S_2	S_1	S_2
5	1,915	1,905	1,659	1,662
10	1,928	1,923	1,654	1,656
15	1,935	1,932	1,652	1,653
20	1,939	1,937	1,651	1,652
25	1,942	1,940	1,650	1,651
30	1,944	1,943	1,650	1,650
40	1,947	1,946	1,649	1,649
50	1,949	1,948	1,648	1,648
∞	1,960	1,960	1,645	1,645

При $n > 20$ справедливо приближение $S_j(\alpha) \approx u_\alpha$ (u_α — α-квантиль стандартного нормального распределения). Статистика S_1^* имеет наибольшую эффективность для распределений симметричного типа, S_2^* — для распределений, скошенных вправо. Критерии S_1^* и S_2^* не уступают по мощности критерию Хсу (см. раздел 4.3.4) [505], но не требует знания параметров положения.

Задача 279. *Проверить гипотезу изменения дисперсии ряда для данных задачи* 278 *критериями* S_1^* *и* S_2^*.

Ранги выборочной последовательности составляют ряд

$$\begin{array}{lcccccccccc} x_i: & 2 & 1 & 11 & 21 & 3 & 8 & 6 & 23 & 38 & 37; \\ R_i: & 2 & 1 & 6 & 7 & 3 & 5 & 4 & 8 & 10 & 9. \end{array}$$

Находим метки $a_{1,n}(i) = u^2_{\frac{i}{n+1}}$, используя аппроксимацию 15 (см. раздел 1.1.1)

$$u_{\frac{i}{n+1}} = 4{,}91 \cdot \left[\left(\frac{i}{n+1}\right)^{0,14} - \left(1 - \frac{i}{n+1}\right)^{0,14}\right].$$

Вычисляем последовательно метки a_{1n} и a_{2n}, сводя результаты в таблицу:

n	$a_{1,10}(i)$	$a_{2,10}(i)$	i	$a_{1,10}(i)$	$a_{2,10}(i)$
1	1,7822	0,1000	6	0,0129	0,8456
2	0,8217	0,2111	7	0,1206	1,0956
3	0,3630	0,3361	8	0,3630	1,4290
4	0,1206	0,4790	9	0,8217	1,9290
5	0,0129	0,6456	10	1,7822	2,9290

Вычисляем статистики критериев

$$S_1 = \sum_{i=1}^{10} i \cdot a_{1,10}(R_i) = 1 \cdot a_{1,10}(2) + 2 \cdot a_{1,10}(1) + \ldots + 10 \cdot a_{1,10}(10) =$$
$$= 0{,}8217 + 2 \cdot 1{,}7822 + \ldots + 10 \cdot 0{,}8217 = 34{,}804;$$

$$S_2 = \sum_{i=1}^{10} i \cdot a_{2,10}(R_i) = 1 \cdot 0{,}2111 + 2 \cdot 0{,}1 + \ldots + 10 \cdot 1{,}929 = 73{,}295.$$

Далее находим

$$\mathbf{M}(S_1) = \frac{11}{2} \cdot \sum_{i=1}^{10} a_{1,10}(i) = \frac{11 \cdot 6{,}201}{2} = 34{,}104; \quad \mathbf{M}(S_2) = \frac{10 \cdot 11}{2} = 55;$$

$$\mathbf{D}(S_1) = \frac{10 \cdot 11}{12} \cdot \sum_{i=1}^{10} a_{1,10}^2(i) - \frac{34{,}102^2}{33} = 38{,}05 \quad (\sqrt{\mathbf{D}(S)} = 6{,}168);$$

$$\mathbf{D}(S_2) = \frac{10 \cdot 11}{12} \cdot \left(10 - \sum_{j=1}^{10} \frac{1}{j}\right) = 64{,}818 \quad (\sqrt{\mathbf{D}\,(S_2)} = 8{,}051);$$

$$|S_1^*| = \frac{|S - \mathbf{M}(S_1)|}{\sqrt{\mathbf{D}(S_1)}} = \frac{|34{,}8 - 34{,}1|}{6{,}168} = 0{,}113; \quad |S_2^*| = \frac{|S_2 - \mathbf{M}(S_2)|}{\sqrt{\mathbf{D}(S_2)}} = \frac{|173{,}29 - 55|}{8{,}051} = 2{,}271.$$

Из табл. 174 имеем $S_1(0{,}95) = 1{,}928$ и $S_2(0{,}95) = 1{,}923$.

Так как $|S_1^*| = 0{,}113 < S_1(0{,}95) = 1{,}928$ и $|S_2^*| = 2{,}271 > S_2(0{,}95) = 1{,}923$, нулевая гипотеза отклоняется критерием S_2 и не отклоняется критерием S_1.

4.3.6. Сериальный критерий случайности

Предположим, имеются две выборки случайных величин x и y. Частичная последовательность элементов одной из выборок в упорядоченной по возрастанию объединенной выборке, ограниченная с обеих сторон элементами другой выборки (на границах последовательности — с одной стороны), называется серией.

Например, последовательность

$$\underline{x, x},\ \underline{y, y},\ \underline{x},\ \underline{y},\ \underline{x, x, x},\ \underline{y, y}$$

содержит 6 серий (3 серии элементов y и 3 серии элементов x).

Последовательность $\underline{x, x, x}$, $\underline{y, y, y}$ содержит две серии (одну элементов x и одну элементов y). Количество и структура серий характеризует случайность появления последовательности против альтернативы тренда, т. е. наличия закономерности в расположении элементов одной выборки.

Например, малое число серий будет указывать на тенденцию объединения элементов одной выборки в группы, т. е. на неслучайный характер их расположения в общей последовательности.

Большинство сериальных критериев, основанных на числе и структуре серий, обладает невысокой мощностью, но требует минимальных расчетов.

4.3.6.1. Критерий Вальда–Волфовитца

Имеется выборка значений случайной величины x в порядке их появления; $\tilde{x}$ — выборочная медиана. Значения $x_i \geqslant \tilde{x}$ обозначаем символом a, а значения $x_i < \tilde{x}$ символом b. Статистикой критерия является N — общее число полученных серий элементов a и b.

Гипотеза случайности ряда принимается с вероятностью α, если $n_1(\alpha) < N < n_2(\alpha)$; в ином случае она отклоняется в пользу альтернативы неслучайности ряда.

Таблица 175

Критические значения $N_1(\alpha)$ и $N_2(\alpha)$ сериального критерия Вальда–Волфовитца [25]

m	n	Доверительная вероятность α				m	n	Доверительная вероятность α			
		0,90		0,95				0,90		0,95	
		N_1	N_2	N_1	N_2			N_1	N_2	N_1	N_2
2	2	1	5	1	5	4	15	3	10	3	10
2	3	1	6	1	6	4	16	4	10	3	10
2	4	1	6	1	6	4	17	4	10	3	10
2	5	1	6	1	6	4	18	4	10	3	10
2	6	1	6	1	6	4	19	4	10	3	10
2	7	1	6	1	6	4	20	4	10	3	10
2	8	1	6	1	6	5	5	2	10	2	10
2	9	1	6	1	6	5	6	3	10	2	11
2	10	1	6	1	6	5	7	3	11	2	11
2	11	1	6	1	6	5	8	3	11	3	12
2	12	2	6	1	6	5	9	3	12	3	12
2	13	2	6	1	6	5	10	3	12	3	12
2	14	2	6	1	6	5	11	4	12	3	12
2	15	2	6	1	6	5	12	4	12	3	12
2	16	2	6	1	6	5	13	4	12	3	12
2	17	2	6	1	6	5	14	4	12	3	12
2	18	2	6	1	6	5	15	4	12	4	12
2	19	2	6	2	6	5	16	4	12	4	12
3	3	1	7	1	7	5	17	4	12	4	12
3	4	1	8	1	8	5	18	5	12	4	12
3	5	1	8	1	8	5	19	5	12	4	12
3	6	2	8	1	8	5	20	5	12	4	12
3	7	2	8	1	8	6	6	3	11	2	12
3	8	2	8	1	8	6	7	3	12	3	12
3	9	2	8	2	8	6	8	3	12	3	13
3	10	2	8	2	8	6	9	4	13	3	13
3	11	2	8	2	8	6	10	4	13	3	13
3	12	2	8	2	8	6	11	4	13	4	14
3	13	2	8	2	8	6	12	4	13	4	14
3	14	2	8	2	8	6	13	5	14	4	14
3	15	3	8	2	8	6	14	5	14	4	14
3	16	3	8	2	8	6	15	5	14	4	14
3	17	3	8	2	8	6	16	5	14	4	14
3	18	3	8	2	8	6	17	5	14	4	14
3	19	3	8	2	8	6	18	5	14	5	14
3	20	3	8	2	8	6	19	6	14	5	14
4	4	1	9	1	9	6	20	6	14	5	14
4	5	2	9	1	9	7	7	3	13	3	13
4	6	2	9	2	10	7	8	4	13	3	14
4	7	2	10	2	10	7	9	4	14	4	14
4	8	3	10	2	10	7	10	4	14	4	15
4	9	3	10	2	10	7	11	4	14	4	15
4	10	3	10	2	10	7	12	4	14	4	15
4	11	3	10	2	10	7	13	15	15	5	16
4	12	3	10	3	10	7	14	15	15	5	16
4	13	3	10	3	10	7	15	15	15	5	16
4	14	3	10	3	10	7	16	16	16	5	16

Продолжение таблицы 175

m	n	Доверительная вероятность α				m	n	Доверительная вероятность α			
		0,90		0,95				0,90		0,95	
		N_1	N_2	N_1	N_2			N_1	N_2	N_1	N_2
7	17	16	16	5	16	11	20	9	21	8	22
7	18	16	16	5	16	12	12	7	19	7	19
7	19	16	16	6	16	12	13	8	19	7	20
7	20	16	16	6	16	12	14	8	20	7	21
8	8	14	14	4	14	12	15	8	20	8	21
8	9	14	14	4	15	12	16	9	21	8	22
8	10	15	15	4	15	12	17	9	21	8	22
8	11	15	15	5	16	12	18	9	21	8	22
8	12	16	16	5	16	12	19	10	22	9	23
8	13	16	16	5	17	12	20	10	22	9	23
8	14	16	16	5	17	13	13	8	20	7	21
8	15	16	16	5	17	13	14	9	20	8	21
8	16	6	17	6	17	13	15	9	21	8	22
8	17	7	17	6	18	13	16	9	21	8	22
8	18	7	17	6	18	13	17	10	22	9	23
8	19	7	17	6	18	13	18	10	22	9	23
8	20	7	17	6	18	13	19	10	23	9	24
9	9	5	15	4	16	13	20	10	23	10	24
9	10	5	16	5	16	14	14	9	21	8	22
9	11	6	16	5	17	14	15	9	22	8	23
9	12	6	16	5	17	14	16	10	22	9	23
9	13	6	17	6	18	14	17	10	23	9	24
9	14	7	17	6	18	14	18	10	23	9	24
9	15	7	18	6	18	14	19	11	23	10	24
9	16	7	18	6	18	14	20	11	24	10	25
9	17	7	18	7	19	15	15	10	22	9	23
9	18	8	18	7	19	15	16	10	23	9	24
9	19	8	18	7	19	15	17	11	23	10	24
9	20	8	18	7	19	15	18	11	24	10	25
10	10	6	16	5	17	15	19	11	24	10	25
10	11	6	17	5	18	15	20	12	25	11	26
10	12	7	17	6	18	16	16	11	23	10	24
10	13	7	18	6	19	16	17	11	24	10	25
10	14	7	18	6	19	16	18	11	25	10	26
10	15	7	18	7	19	16	19	12	25	11	26
10	16	8	19	7	20	16	20	12	25	11	26
10	17	8	19	7	20	17	17	11	25	10	26
10	18	8	19	7	20	17	18	12	25	11	26
10	19	8	20	8	20	17	19	12	26	11	27
10	20	9	20	8	20	17	20	13	26	11	27
11	11	7	17	6	18	18	18	12	26	11	27
11	12	7	18	6	19	18	19	13	26	12	27
11	13	7	19	6	19	18	20	13	27	12	28
11	14	8	19	7	20	19	19	13	27	12	28
11	15	8	19	7	20	19	20	13	27	12	29
11	16	8	20	7	21	20	20	14	28	13	29
11	17	9	20	8	21	21	21	15	28	14	29
11	18	9	20	8	21	22	22	16	29	14	31
11	19	9	21	8	22	23	23	16	31	15	32

Окончание таблицы 175

m	n	Доверительная вероятность α				m	n	Доверительная вероятность α			
		0,90		0,95				0,90		0,95	
		N_1	N_2	N_1	N_2			N_1	N_2	N_1	N_2
24	24	17	32	16	33	38	38	30	47	28	49
25	25	18	33	17	34	39	39	30	49	29	50
26	26	19	34	18	35	40	40	31	50	30	51
27	27	20	35	19	36	41	41	32	51	31	52
28	28	21	36	19	38	42	42	33	52	31	54
29	29	22	37	20	39	43	43	34	53	32	55
30	30	22	39	21	40	44	44	34	54	33	56
31	31	23	40	22	41	45	45	36	55	34	57
32	32	24	41	23	42	46	46	37	56	35	58
33	33	25	42	24	43	47	47	38	57	36	59
34	34	26	43	24	45	48	48	38	59	37	60
35	35	27	44	25	46	49	49	39	60	38	61
36	36	28	45	26	47	50	50	40	61	38	63
37	37	29	46	27	48						

Критические значения $n_1(\alpha)$ и $n_2(\alpha)$ приведены в табл. 175. Если m и n — соответственно количества элементов a и b в последовательности, то при $m, n > 20$ справедлива аппроксимация

$$N^* = \frac{N - \left(\dfrac{2mn}{m+n} + 1\right)}{\sqrt{\dfrac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)}}},$$

имеющая стандартное нормальное распределение. Тогда нулевая гипотеза отклоняется с достоверностью α, если $|N^*| > u_{\frac{1+\alpha}{2}}$.

Задача 280. *Для ряда значений задачи 278 проверить гипотезу случайности ряда наблюдений критерием Вальда–Волфовитца при доверительной вероятности* $\alpha = 0{,}90$.

Имеем медиану ряда $\tilde{x} = \dfrac{8+11}{2} = 9{,}5$. Обозначим символом a значения $x_i < \tilde{x}$ и символом b значения $x_i > \tilde{x}$. Получаем ряд символов

$$\underline{a, a},\ \underline{b, b},\ \underline{a, a, a},\ \underline{b, b, b}.$$

В нашем случае имеем $m = 5$ элементов a и $n = 5$ элементов b. Всего имеем $N = 4$ серии элементов — две серии элементов a и две серии элементов b.

Из табл. 175 для $m = n = 5$ и $\alpha = 0{,}90$ находим $n_1 = 2$ и $n_2 = 10$. Таким образом, $n_1 = 2 < N = 4 < n_2 = 10$, и ряд признается случайным. Следует отметить малую мощность таких критериев (их рекомендуется применять при $m, n > 50$). Для нормальной аппроксимации имеем

$$\mathbf{M}(N) = \frac{2 \cdot 5 \cdot 5}{10} + 1 = 6; \quad \mathbf{D}(N) = \frac{2 \cdot 5 \cdot 5 \cdot (2 \cdot 5 \cdot 5 - 5 - 5)}{(5+5)^2 \cdot (5+5-1)} = 2{,}222 \quad (\sqrt{\mathbf{D}(N)} = 1{,}491);$$

$$|N^*| = \frac{|4-6|}{1{,}491} = 1{,}3.$$

что меньше $u_{\frac{1+\alpha}{2}} = u_{0,95} = 1{,}645$; следовательно, гипотеза принимается.

4.3.6.2. Критерий Рамачандрана–Ранганатана

В отличие от критерия Вальда–Волфовитца (см. раздел 4.3.6.1) настоящий критерий учитывает не только количество, но и длины серий (количества элементов в сериях). Статистика критерия имеет вид

$$R = \sum_j j^2 n_j,$$

где j — длина серии, n — объем выборки; n_j — количество серий длины j.

Гипотеза случайности не отклоняется с вероятностью α при $R < R(\alpha)$. Критические значения $R(\alpha)$ приведены в табл. 176. Критерий обладает большей мощностью по сравнению с N-критерием Вальда–Волфовитца.

Таблица 176

Критические значения $R(\alpha)$ критерия Рамачандрана–Ранганатана [14]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
6	18	18	18	20	68	78	100
8	26	32	32	22	76	86	108
10	34	38	50	24	84	94	118
12	38	44	58	26	92	102	128
14	46	52	68	28	98	110	136
16	54	60	80	30	106	118	146
18	68	78	90				

Задача 281. *Для данных задачи* 278 *проверить гипотезу случайности ряда критерием Рамачандрана–Ранганатана.*

Имеем две серии элементов длины $j = 2$ ($n_2 = 2$) и две серии элементов длины $j = 3$ ($n_3 = 2$). Тогда $R = \sum_j j^2 \cdot n_j = 4 \cdot 2 + 9 \cdot 2 = 26$.

Из табл. 176 для $n = 10$ и $\alpha = 0{,}95$ имеем $R(0,95) = 38$.

Так как $R = 26 < R(0{,}95) = 38$, ряд значений x_i признается случайным, а альтернатива тренда отклоняется.

4.3.6.3. Сериальный критерий Шахнесси

Критерий Шахнесси [510] является множественным аналогом критерия Вальда–Волфовитца (см. раздел 4.3.6.1). Если в критерии Вальда–Волфовитца рассматривается количество серий элементов двух „сортов", то критерий Шахнесси предполагает анализ серий элементов k „сортов" ($k \geqslant 2$). Это делает его более эффективным, так как позволяет противостоять большему количеству альтернатив (сдвиг, колебания, изменения в определенной точке).

Статистикой критерия остается, как и ранее, $\tilde{N}$ — общее количество серий элементов. Если количество серий

$$\tilde{N} < N_\alpha(n_i, k)\,,$$

то гипотеза случайности отклоняется с вероятностью α (здесь $N_\alpha(n_i, k)$ — критические значения, приведенные в табл. 177). В табл. 177 приняты следующие обозначения: n_i ($i = 1, \ldots, k$) — количество элементов i-го „сорта", k — количество „сортов" элементов, составляющих ряд.

Таблица 177

Критические значения $N_\alpha(n_i, k)$ критерия Шахнесси для $k = 2 \div 6$ [510]

k	n_1	n_2	α		k	n_1	n_2	α		k	n_1	n_2	α	
			0,95	0,90				0,95	0,90				0,95	0,90
2	4	4	2	2	2	10	11	7	8	2	16	17	12	13
2	4	5	2	3	2	11	11	7	8	2	17	18	13	14
2	5	5	3	3	2	11	12	8	9	2	18	19	14	15
2	6	6	3	4	2	12	12	8	9	2	19	20	14	16
2	6	7	4	4	2	12	13	9	9	2	20	20	15	16
2	7	7	4	5	2	13	13	9	10	2	21	21	16	17
2	8	8	5	5	2	13	14	9	10	2	22	22	17	18
2	8	9	5	6	2	14	14	10	11	2	23	23	17	19
2	9	9	6	6	2	15	15	11	12	2	24	24	18	20
2	9	10	6	7	2	16	16	11	12	2	25	25	19	21

k	n_1	n_2	n_3	α		k	n_1	n_2	n_3	α	
				0,95	0,90					0,95	0,90
3	2	3	3	3	4	3	7	7	77	11	11
3	3	3	3	4	4	3	7	8	8	11	12
3	3	3	4	4	5	3	7	8	8	12	13
3	3	4	4	5	5	3	8	8	8	12	13
3	4	4	4	5	6	3	8	9	9	13	14
3	4	4	5	6	7	3	8	9	9	13	14
3	4	5	5	7	7	3	9	9	9	14	15
3	5	5	5	7	8	3	9	10	10	15	16
3	5	5	6	8	8	3	10	10	10	16	17
3	5	6	6	8	9	3	10	11	11	16	17
3	6	6	6	9	10	3	10	11	11	17	18
3	6	6	7	9	10	3	11	11	11	18	19
3	6	7	7	10	11	3	11	12	12	19	20

k	n_1	n_2	n_3	n_4	α		k	n_1	n_2	n_3	n_4	α	
					0,95	0,90						0,95	0,90
4	2	2	2	2	4	5	4	5	5	5	6	13	13
4	2	2	2	3	5	5	4	5	5	6	6	13	14
4	2	2	3	3	5	6	4	5	6	6	6	14	15
4	2	3	3	3	6	7	4	6	6	6	6	15	15
4	3	3	3	3	7	7	4	6	6	6	7	15	16
4	3	3	3	4	7	8	4	6	6	7	7	16	17
4	3	3	4	4	8	9	4	6	7	7	7	17	17
4	3	4	4	4	9	9	4	7	7	7	7	17	18
4	4	4	4	4	9	10	4	7	7	7	8	18	19
4	4	4	4	5	10	11	4	7	7	8	8	19	20
4	4	4	5	5	11	11	4	7	8	8	8	19	20
4	4	5	5	5	11	12	4	8	8	8	8	20	21
4	5	5	5	5	12	13							

k	n_1	n_2	n_3	n_4	n_5	α		k	n_1	n_2	n_3	n_4	n_5	α	
						0,95	0,90							0,95	0,90
5	2	2	2	2	2	6	7	5	3	3	3	3	3	10	10
5	2	2	2	2	3	7	7	5	3	3	3	3	4	10	11
5	2	2	2	3	3	8	8	5	3	3	3	4	4	11	12
5	2	2	3	3	3	8	9	5	3	3	4	4	4	12	12
5	2	3	3	3	3	9	9	5	3	4	4	4	4	12	13

Окончание таблицы 177

k	n_1	n_2	n_3	n_4	n_5	α		k	n_1	n_2	n_3	n_4	n_5	α	
						0,95	0,90							0,95	0,90
5	4	4	4	4	4	13	14	5	5	5	6	6	6	19	20
5	4	4	4	4	5	14	15	5	5	6	6	6	6	20	21
5	4	4	4	5	5	15	15	5	6	6	6	6	6	21	21
5	4	4	5	5	5	15	16	5	6	6	6	6	7	21	22
5	4	5	5	5	5	16	17	5	6	6	6	7	7	22	23
5	5	5	5	5	5	17	18	5	6	6	7	7	7	23	24
5	5	5	5	5	6	18	18	5	6	7	7	7	7	23	24
5	5	5	5	6	6	18	19	5	7	7	7	7	7	24	25

Задача 282. *Дан ряд наблюдений над случайной величиной*

$$x_i\colon 12,\ 8,\ 6,\ 0,\ -4,\ -3,\ 1,\ -2,\ -6,\ -10,\ 8,\ 4,\ 2,\ -1,\ 5,\ 15,\ 21,\ 32.$$

Необходимо проверить гипотезу случайности ряда критерием Шахнесси при доверительной вероятности $\alpha = 0{,}95$.

Элементам, для которых $x_i < 0$, присвоим код a; элементам, у которых $0 \leqslant x_i \leqslant 7$, присвоим код b, а элементам, для которых $x_i > 7$, присвоим код c. Тогда получаем ряд

$$\underline{c,c},\ \underline{b,b},\ \underline{a,a},\ \underline{b},\ \underline{a,a,a},\ \underline{c},\ \underline{b,b},\ \underline{a},\ \underline{b},\ \underline{c,c}.$$

Видим, что общее количество серий равно 10 (3 серии элементов c, 3 серии элементов a и 4 серии элементов b). Количества элементов равны $n_1 = 6$ (a), $n_2 = 6$ (b) и $n_3 = 6$ (c). В табл. 177 для $k = 3$, $n_1 = 6$, $n_2 = 6$ и $n_3 = 6$ при $\alpha = 0{,}95$ находим $n_\alpha = 9$.

Так как $n = 10 > n_\alpha = 9$, гипотеза случайности не отклоняется.

4.3.6.4. Критерий Олмстеда

Олмстедом [509] рассмотрена серия критериев случайности, так же, как и в случае критерия Рамачандрана–Ранганатана (см. раздел 4.3.6.2) учитывающих длины серий. В критериях Олмстеда рассматриваются экстремальные длины серий одного вида, вероятность появления которых связывается с возможным присутствием тренда в исследуемых рядах. Олмстедом предложено четыре варианта критерия: наибольшая длина l_1 серии, лежащей по какую-либо одну сторону от медианы; наибольшая длина l_2 серии, лежащей по одну (заранее выбранную) сторону от медианы; кратчайшая l_3 из обеих наибольших длин серий, лежащих по разные стороны от медианы; кратчайшая l_4 из обеих наибольших длин серий, лежащих по разные стороны от точки раздела, максимизирующей l_4.

Во всех вариантах для заданных l_i $(i = 1, 2, 3, 4)$ на уровне значимости α статистикой критерия Олмстеда является наименьший объем выборки n. Гипотеза случайности отклоняется при $n < n_\alpha(l_i)$. Критические значения $n_\alpha(l_i)$ приведены в табл. 178.

Задача 283. *Для ряда наблюдений*

$$x_i : 13,\ 8,\ 7,\ 4,\ 10,\ 17,\ 21,\ 34,\ 48,\ 1,\ 0,\ 12,\ 10,\ 4,\ 16,\ 17,\ 11,\ 0,\ 1,\ 3,\ 12,\ 54,\ 16,\ -1$$

($n = 24$) *проверить гипотезу случайности критерием Олмстеда на уровне значимости* $\alpha = 0{,}1$.

Находим медиану ряда $\tilde{x} = 11$. Для значений ряда $x_i \geqslant \tilde{x}$ примем индекс a, а для значений $x_i < \tilde{x}$ — индекс b. Получаем ряд

$$\underline{a},\ \underline{b,b,b,b},\ \underline{a,a,a,a},\ \underline{b,b},\ \underline{a},\ \underline{b,b},\ \underline{a,a,a},\ \underline{b,b,b},\ \underline{a,a,a},\ \underline{b},$$

в котором 5 серий элементов a и 5 серий элементов b, причем $l_1 = 4$, $l_2 = 4$, $l_3 = 4$. Из табл. 178 для $\alpha = 0{,}1$, $l_1 = l_2 = l_3 = 4$ находим $n(l_1) = 8$, $n(l_2) = 10$ и $n(l_3) = 14$. Так как $n = 24 > n(l_1)$, $n(l_2)$, $n(l_3)$, то всеми критериями гипотеза случайности не отклоняется. Легко убедиться, что и последний критерий (l_4) также отклоняет гипотезу тренда.

Таблица 178

Критические значения $n_\alpha(l_i)$ критерия случайности Олмстеда [24]

l_i	Уровень значимости α							
	0,10				0,01			
	$n(l_1)$	$n(l_2)$	$n(l_3)$	$n(l_4)$	$n(l_1)$	$n(l_2)$	$n(l_3)$	$n(l_4)$
2	4	4	4	4	4	4	4	4
3	6	6	8	8	6	6	6	6
4	8	10	14	12	8	8	8	8
5	14	16	26	18	10	10	12	12
6	20	26	50	34	12	14	20	16
7	32	44	98	58	16	18	34	24
8	52	78	194	108	22	26	62	38
9	86	142	390	204	32	38	116	66
10	150	256	782	400	42	56	216	118
11	262	480	1182	790	62	86	446	228
12	500	930	2360	1568	94	140	884	444
13	962	1838	4720	3130	156	234	1762	878
14	1876	3630	9450	6220	254	410	3510	1750
15	3670	7160	18900	12490	418	748	6990	3480
16	7330	14190	37800	25000	766	1446	13930	6790
17	14090	28100	75600	49900	1472	2830	27900	13865
18	27900	56100	151200	99900	2860	5530	55500	27700
19	555000	117300	30200	199800	5570	10860	111000	55400
20	1111000	235000	60500	400000	10860	21500	222000	110800

4.3.6.5. Критерий числа серий знаков первых разностей

Для выборки $x_1, x_2, \ldots, x_n$ вычисляем $(n-1)$ значений вида

$$z_i = \begin{cases} -1, & \text{если } x_{i+1} < x_i; \\ +1, & \text{если } x_{i+1} > x_i; \\ 0, & \text{если } x_{i+1} = x_i. \end{cases}$$

В ряду значений z_i фиксируем количество серий R, которое и является статистикой рассматриваемого критерия.

Гипотеза случайности ряда не отклоняется при $R_1(\alpha) < R < R_2(\alpha)$. Критические значения $R_1(\alpha)$ и $R_2(\alpha)$ приведены в табл. 179. При $n > 30$ распределение R удовлетворительно аппроксимируется нормальным распределением со средним $\mathbf{M}(R)$ и дисперсией $\mathbf{D}(R)$, где $\mathbf{M}(R) = \dfrac{2n-1}{3}$; $\mathbf{D}(R) = \dfrac{16n-29}{90}$. Тогда нулевая гипотеза проверяется критерием $|R^*| = \dfrac{|R - \mathbf{M}(R)|}{\sqrt{\mathbf{D}(R)}}$ и при $|R^*| \geqslant u_{\frac{1+\alpha}{2}}$ гипотеза случайности отклоняется.

Хальд [51], Уоллис и Мур [413] дополнили критерий числа серий рассмотрением длин серий знаков + и −. Вероятность появления серий длины l при справедливости гипотезы случайности ряда равна

$$p(l) = \frac{2}{n}\left[\frac{n-1}{(l+1)!} - 2\frac{n-l-1}{(l+2)!} + \frac{n-l-2}{(l+3)!}\right].$$

Таблица 179

Критические значения $R_1(\alpha)$ и $R_2(\alpha)$ критерия числа серий знаков первых разностей [57]

n	Доверительная вероятность α				n	Доверительная вероятность α			
	0,95		0,90			0,95		0,90	
	R_1	R_2	R_1	R_2		R_1	R_2	R_1	R_2
5	1		1		30	14	15	15	24
6	1		1		32	15	26	16	26
7	2		2		34	17	28	17	27
8	2		2		36	18	29	19	29
9	2		2		38	19	31	20	30
10	2		3		40	20	32	21	32
11	4	10	4	10	42	21	34	22	33
12	4	11	4	11	44	23	35	23	34
13	5	12	5	12	46	24	37	25	36
14	5	13	6	12	48	25	38	26	37
15	6	14	6	13	50	26	40	27	39
16	6	14	7	14	52	27	41	28	40
17	7	15	7	15	54	29	43	30	42
18	7	16	8	15	56	30	44	31	43
19	8	17	8	16	58	31	45	32	44
20	8	17	9	17	60	32	47	33	46
21	9	18	10	18	62	34	48	35	47
22	10	19	10	18	64	35	50	36	49
23	10	20	11	19	66	36	51	37	50
24	11	20	11	20	68	37	53	38	52
25	11	21	12	21	70	38	54	40	53
26	12	22	13	21	80	45	61	46	60
27	13	23	13	22	90	51	68	52	67
28	13	23	14	23	100	57	75	58	74
29	14	24	14	24					

Критерий с учетом длин серий основан на статистике

$$\chi^2 = \sum_l \frac{[n_l - np(l)]^2}{np(l)},$$

где суммирование ведется для $l = 1, 2$ и $\geqslant 3$ (n_l — количество серий длины l, $p(l)$ — вероятность появления серии длины l). В частности,

$$p(1) = \frac{2}{n}\left(\frac{n-1}{2} - 2\frac{n-2}{6} + \frac{n-3}{24}\right) = \frac{5n+1}{12n};$$

$$p(2) = \frac{2}{n}\left(\frac{n-2}{6} - 2\frac{n-3}{24} + \frac{n-4}{120}\right) = \frac{11n-14}{60n}; \quad p(3) = 1 - p(1) - p(2) = \frac{7-15n}{12n}.$$

Правило проверки нулевой гипотезы формулируется следующим образом:

— при $\chi^2 \geqslant 6{,}3$ нулевая гипотеза отклоняется с вероятностью α, если $\chi^2 > \chi^2_\alpha(2{,}5)$, где $\chi^2_\alpha(2{,}5)$ — α-квантиль распределения хи-квадрат с $f = 2{,}5$ степенями свободы;

— при $\chi^2 < 6{,}3$ нулевая гипотеза отклоняется, если $\chi^2 > \dfrac{7}{6}\chi^2(2)$.

Задача 284. *Проверить гипотезу случайности ряда из задачи 283 критерием числа серий знаков первых разностей при доверительной вероятности* $\alpha = 0{,}95$.

Ряд знаков первых разностей имеет вид ($n = 24$):

$$- - - + + + + + - - + - - + + - - + + + + - -$$

Из него получаем число серий $R = 9$ (5 серий „$-$“ и 4 серии „$+$“).

Из табл. 179 для $n = 24$ и $\alpha = 0{,}95$ имеем $R_1 = 11$ и $R_2 = 20$.

Так как $R = 8 < R_1 = 11$, гипотеза случайности выборки x отклоняется.

Для нормальной аппроксимации имеем

$$\mathbf{M}(R) = \frac{2 \cdot 24 - 1}{3} = 15{,}67; \quad \mathbf{D}(R) = \frac{16 \cdot 24 - 29}{90} = 3{,}944 \quad (\sqrt{\mathbf{D}(R)} = 1{,}986);$$

$$|R^*| = \frac{|11 - 15{,}7|}{1{,}986} = 2{,}366.$$

Так как $|R^*| = 2{,}366 > u_{0,975} = 1{,}96$, то и этот критерий отклоняет гипотезу случайности ряда.

Рассмотрим теперь критерий с учетом длин серий. Вычисляем

$$p\,(1) = \frac{5 \cdot 24 + 1}{12 \cdot 24} = 0{,}42; \quad p\,(2) = \frac{11 \cdot 24 - 14}{60 \cdot 24} = 0{,}1736; \quad p\,(\geqslant 3) = 1 - 0{,}42 - 0{,}1736 = 0{,}4064.$$

В нашем случае имеются $n_1 = 1$ серия длины $l = 1$, $n_2 = 5$ серий длины $l = 2$ и $n_3 = 3$ серии длины $l = 3$. Находим

$$\chi^2 = \sum_l \frac{[n_l - n \cdot p\,(l)]^2}{n \cdot p\,(l)} = \frac{(1 - 24 \cdot 0,42)^2}{24 \cdot 0,42} + \frac{(5 - 24 \cdot 0,1736)^2}{24 \cdot 0,1736} + \frac{(3 - 24 \cdot 0,4046)^2}{24 \cdot 0,4046} = 12,983.$$

Так как $\chi^2 > 6{,}3$, то используем критическое значение $\chi^2_{0,95}(2{,}5) = 6{,}903$.

Из того, что $\chi^2 = 12{,}983 > \chi^2_{0,95}(2{,}5) = 6{,}9$, следует, что и этот критерий отклоняет гипотезу случайности.

4.3.7. Критерий инверсий

Если в выборке значений $x_1, x_2, \ldots, x_n$, записанных в порядке их появления, за некоторым значением x_i следует меньшее по величине (т. е. $x_i > x_j$, где $i + 1 \leqslant j \leqslant n$), то имеет место инверсия. Общее число инверсий I в выборке является статистикой критерия случайности полученных значений x [14]. Иногда рассматривают статистику T, определяемую числом обратных инверсий (когда $x_i < x_j$, $j > i$). Статистика T совпадает со статистикой U Манна–Уитни (см. раздел 4.2.1.1.2.2), и ее критическое значение может быть определено по табл. 143 этого критерия. В качестве третьей меры можно использовать величину $K = T - I$, критические значения которой приведены в работе Кендалла [422].

При $n \geqslant 20$ статистика I распределена приблизительно нормально со средним $\mathbf{M}(I)$ и дисперсией $\mathbf{D}(I)$, где

$$\mathbf{M}(I) = \frac{n(n-1)}{4}; \quad \mathbf{D}(I) = \frac{2n^3 + 3n^2 - 5n}{72}.$$

При $|I^*| = \dfrac{|I - \mathbf{M}(I)|}{\sqrt{\mathbf{D}(I)}} \geqslant u_{\frac{1+\alpha}{2}}$ гипотеза случайности отклоняется с вероятностью α. Точные критические значения $I_1(\alpha)$ и $I_2(\alpha)$ приведены в табл. 180 — гипотеза случайности принимается, если

$$I_1(\alpha) < I < I_2(\alpha)\,.$$

Критерий имеет асимптотическую эффективность $\sqrt[3]{3\pi} \approx 0{,}98$ относительно критерия коэффициента регрессии. Следовательно, по эффективности он превосходит большинство непараметрических критериев для тренда.

Таблица 180

Критические значения $I_1(\alpha)$ и $I_2(\alpha)$ критерия инверсий [14]

n	Доверительная вероятность α				n	Доверительная вероятность α			
	0,90		0,95			0,90		0,95	
	I_1	I_2	I_1	I_2		I_1	I_2	I_1	I_2
10	13	31	11	33	40	319	460	305	474
12	21	44	18	47	50	514	710	495	729
14	30	60	27	63	60	756	1013	731	1038
16	41	78	38	81	70	1045	1369	1014	1400
18	54	98	50	102	80	1382	1777	1344	1815
20	69	120	64	125	90	1766	2238	1721	2283
30	171	263	162	272	100	2198	2751	2145	2804

Задача 285. *Проверить гипотезу случайности ряда*

x_i: 13, 8, 7, 4, 10, 17, 21, 34, 48, 1, 12, 9, 3, 18, −1, −3, 49, 50, 0, 14

критерием инверсий при доверительной вероятности $\alpha = 0{,}95$.

Определяем количества инверсий I для различных значений x_i, результаты сведем в таблицу:

i	x_i	I	i	x_i	I	i	x_i	I
1	13	11	8	34	9	15	−1	1
2	8	7	9	48	9	16	−3	0
3	7	6	10	1	3	17	49	2
4	4	5	11	12	5	18	50	2
5	10	6	12	9	4	19	0	0
6	17	8	13	3	3			
7	21	9	14	18	4			

Сумма инверсий равна $I = 11+7+\ldots+2+0 = 94$. Из табл. 180 для $n = 20$ и $\alpha = 0{,}95$ имеем $I_1 = 64$ и $I_2 = 272$.

Так как $I_1 = 64 < I = 94 < I_2 = 272$, гипотеза случайности не отклоняется.

Используем теперь нормальную аппроксимацию, для чего находим

$$\mathbf{M}(I) = \frac{20 \cdot 19}{4} = 95; \quad \mathbf{D}(I) = \frac{2 \cdot 20^3 + 3 \cdot 20^2 - 5 \cdot 20}{72} = 237{,}5 \quad (\sqrt{\mathbf{D}(I)} = 15{,}411);$$

$$|I^*| = \frac{|94 - 95|}{15{,}411} = 0{,}065.$$

Так как $|I^*| = 0{,}065 \ll u_{0{,}975} = 1{,}96$, гипотеза случайности также уверенно не отклоняется.

4.3.8. Критерий автокорреляции

Если выборка значений x случайна, то значение каждого ее элемента не должно зависеть от величины предшествующего и последующих членов. Для проверки этой независимости используется статистика [36]

$$r_{1,n} = \frac{n \sum_{i=1}^{n-1} x_i x_{i+1} - \left(\sum_{i=1}^{n} x_i\right)^2 + n x_1 x_n}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2},$$

являющаяся коэффициентом корреляции первого порядка между элементами первичной выборки $(x_1, x_2, \ldots, x_n)$ и элементами выборки, полученной из нее сдвигом на одну единицу $(x_2, x_3, \ldots, x_n, x_1)$.

Гипотеза случайности ряда принимается с вероятностью α при $r'_{1,n}(\alpha) < r_{1,n} < r''_{1,n}(\alpha)$. Критические значения коэффициентов автокорреляции $r'_{1,n}(\alpha)$ и $r''_{1,n}(\alpha)$ приведены в табл. 181.

Таблица 181

Критические значения $r'_{1,n}(\alpha)$ и $r''_{1,n}(\alpha)$ критерия автокорреляции [22, 511]

n	Доверительная вероятность α				n	Доверительная вероятность α			
	0,95		0,99			0,95		0,99	
	$r'_{1,n}$	$r''_{1,n}$	$r'_{1,n}$	$r''_{1,n}$		$r'_{1,n}$	$r''_{1,n}$	$r'_{1,n}$	$r''_{1,n}$
5	−0,753	0,253	−0,798	0,297	45	−0,262	0,218	−0,356	0,313
10	−0,564	0,360	−0,705	0,525	50	−0,248	0,209	−0,339	0,300
15	−0,462	0,328	−0,597	0,475	55	−0,236	0,201	−0,324	0,288
20	−0,399	0,299	−0,524	0,432	60	−0,226	0,193	−0,310	0,277
25	−0,356	0,276	−0,473	0,398	65	−0,217	0,186	−0,298	0,267
30	−0,324	0,257	−0,433	0,370	70	−0,209	0,180	−0,287	0,258
35	−0,299	0,242	−0,401	0,347	75	−0,201	0,174	−0,276	0,250

При $n \geqslant 75$ величину $r_{1,n}$ можно считать распределенной асимптотически нормально со средним $\mathbf{M}(r_{1,n})$ и дисперсией $\mathbf{D}(r_{1,n})$, где

$$\mathbf{M}(r_{1,n}) = -\frac{1}{n-1}; \quad \mathbf{D}(r_{,n}) = \frac{n(n-3)}{(n+1)(n-1)^2}.$$

Поэтому критерий случайности может быть записан в форме

$$|r^*_{1,n}| = \frac{|r_{1,n} - \mathbf{M}(r_{,n})|}{\sqrt{\mathbf{D}(r_{,n})}} < u_{\frac{1+\alpha}{2}}.$$

Среди других удачных нормализующих преобразований укажем результаты

— Люнга–Бокса [500]: $r^*_{1,n} = \left\{\dfrac{n(n+2)}{n-1}\right\}^{\frac{1}{2}} r_{1,n}$;

— Морана [551]: $r^*_{1,n}(n-1)^{\frac{1}{2}}\, \dfrac{nr_{1,n}+1}{n-2}$;

— Дюффа–Роя [501]: $r^*_{1,n} = \left[\dfrac{n-1}{n(n-2)}\right]^{\frac{1}{2}} (nr_{1,n}+1)$.

Для более сложных альтернатив (например, периодические колебания ряда) более эффективны различные модификации критериев автокорреляции. Укажем на некоторые из них.

Сумма коэффициентов корреляции первого и второго порядков [513]

$$r_{1,2} = \frac{\sum\limits_{i=1}^{n-1}(x_i-\bar{x})(x_{i+1}-\bar{x}) + \sum\limits_{i=1}^{n-2}(x_i-\bar{x})(x_{i+2}-\bar{x})}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2},$$

имеющая при $n \geqslant 20$ приближенно нормальное распределение со средним $\mathbf{M}(r_{1,2})$ и дисперсией $\mathbf{D}(r_{1,2})$, где

$$\mathbf{M}(r_{1,2}) = -\frac{2n-3}{n(n-1)}; \quad \mathbf{D}(r_{1,2}) = \frac{2n^4 - 13n^3 + 15n + 28n - 34}{n(n-1)^2(n+1)}.$$

Линейная комбинация сериальных коэффициентов

$$r_l = \frac{\sum\limits_{j=1}^{n-1} \frac{1}{j} \sum\limits_{j=1}^{n-j} (x_i - \bar{x})(x_{i+j} - \bar{x})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2},$$

также аппроксимирующаяся нормальным распределением со средним $\mathbf{M}(r_l) = \frac{1}{n} \sum\limits_{i=1}^{n-1} \frac{1}{i(n-1)}$ и дисперсией, приведенной в таблице для различных n:

n	10	16	32	64
$\mathbf{D}(r_l)$	0,041	0,036	0,027	0,018

Во всех случаях при $n \geqslant 20$ справедливо следующее решение: если $\left|r^*_{1,2}\right| = \frac{|r_{1,2} - \mathbf{M}(r_{1,2})|}{\sqrt{\mathbf{D}(r_{1,2})}} \left(|r^*_l| = \frac{|r_l - \mathbf{M}(r_l)|}{\sqrt{\mathbf{D}(r_l)}}\right) > u_{\frac{1+\alpha}{2}}$, то нулевая гипотеза случайности ряда отклоняется.

Задача 286. *Проверить гипотезу случайности ряда значений x, заданных в задаче 285, критерием автокорреляции при доверительной вероятности $\alpha = 0,95$.*

Находим

$$\left(\sum_{i=1}^{20} x_i\right)^2 = 98956; \quad \sum_{i=1}^{19} x_i \cdot x_{i+1} = 5638; \quad \sum_{i=1}^{20} x_i^2 = 10254;$$

$$r_{,20} = \frac{20 \cdot 5638 - 98596 + 20}{20 \cdot 10254 - 98596} = 0{,}167.$$

Из табл. 181 находим $r'_{1,20}(0{,}95) = -0{,}399$ и $r''_{1,20}(0{,}95) = 0{,}299$.

Так как $r'_{1,20}(0{,}95) = -0{,}399 < r_{1,20} = 0{,}167 < r''_{1,20}(0{,}95) = 0{,}299$, гипотеза случайности не отклоняется.

Используем теперь аппроксимацию. Имеем

$$\mathbf{M}(r_{1,20}) = -\frac{1}{19} = -0{,}526; \quad \mathbf{D}(r_{1,20}) = \frac{20 \cdot 17}{21 \cdot 19^2} = 0{,}0448 \quad (\sqrt{\mathbf{D}(r_{1,20})} = 0{,}212);$$

$$\left|r^*_{1,20}\right| = \frac{|0{,}167 + 0{,}0526|}{0{,}212} = 1{,}037.$$

Так как $|r^*_{1,20}| = 1{,}037 < u_{0,975} = 1{,}96$, нулевая гипотеза не отклоняется.

Применим теперь модификацию коэффициента сериальной корреляции. Вычисляем

$$r_{1,2} = \frac{\sum\limits_{i=1}^{19} (x_i - \bar{x}) \cdot (x_{i+1} - \bar{x}) + \sum\limits_{=i1}^{18} (x_i - \bar{x}) \cdot (x_{i+2} - \bar{x})}{\sum\limits_{i=1}^{20} (x_i - \bar{x})^2};$$

$$\bar{x} = 15{,}7; \quad \sum_{i=1}^{20} (x_i - \bar{x}) = 5324{,}3; \quad \sum_{i=1}^{19} (x_i - \bar{x}) \cdot (x_{i+1} - \bar{x}) = 902{,}61;$$

$$\sum_{i=1}^{18} (x_i - \bar{x}) \cdot (x_{i+2} - \bar{x}) = 1442{,}88.$$

Окончательно имеем

$$r_{1,2} = \frac{902{,}61 - 1442{,}88}{5324{,}3} = -0{,}1014; \quad \mathbf{M}(r_{1,2}) = -\frac{2 \cdot 20 \cdot 31}{20 \cdot 19} = 0{,}0973;$$

$$\mathbf{D}(r_{1,2}) = \frac{2\cdot 20^4 - 13\cdot 20^3 + 15\cdot 20^2 + 28\cdot 20 - 34}{20^2\cdot 19\cdot 21} = 0{,}07338 \quad (\sqrt{\mathbf{D}(r_{1,2})} = 0{,}2709);$$

$$\left|r^*_{1,2}\right| = \frac{|-0{,}101 + 0{,}097|}{0{,}0733} = 0{,}054.$$

Так как $|r^*_{1,2}| = 0{,}054 < u_{0,975} = 1{,}96$, гипотеза случайности ряда принимается.

4.3.9. Критерии ранговой корреляции

4.3.9.1. Критерий Вальда–Волфовитца [514]

Пусть R_i — ранг наблюдения x_i в упорядоченном по возрастанию ряду значений $x_1, x_2, \ldots, x_n$. В качестве аналога критерия сериальной корреляции (см. раздел 4.3.8) известен коэффициент ранговой сериальной корреляции Вальда–Волфовитца [512, 514]

$$R = \sum_{i=1}^{n-1}\left(R_i - \frac{n+1}{2}\right)\left(R_{i+1} - \frac{n+1}{2}\right).$$

При $n > 20$ распределение R асимптотически нормально [514, 515] со средним $\mathbf{M}(R)$ и дисперсией $\mathbf{D}(R)$, где

$$\mathbf{M}(R) = 0; \quad \mathbf{D}(R) = \frac{n^2(n+1)(n-3)(5n+6)}{720}.$$

Тогда если $|R^*| = \dfrac{|R|}{\sqrt{\mathbf{D}(R)}} > u_{\frac{1+\alpha}{2}}$, то с вероятностью α гипотеза случайности отклоняется. Для распределений нормального типа асимптотическая эффективность R-критерия $\approx 0{,}91$ по отношению к сериальной корреляции первого порядка (см. раздел 4.3.8), для любых других распределений — не менее 0,86 [512].

Задача 287. *Проверить гипотезу случайности ряда значений, заданных в задаче* 284, *критерием ранговой автокорреляции при доверительной вероятности* $\alpha = 0{,}95$.

Имеем ряд x_i (i — порядковый номер x_i в последовательности, R_i — ранг x_i в последовательности):

i	x_i	R_i	i	x_i	R_i	i	x_i	R_i	i	x_i	R_i
1	13	12	6	17	14	11	12	11	16	−3	2
2	8	8	7	21	16	12	9	9	17	49	19
3	7	7	8	34	17	13	3	5	18	50	20
4	4	6	9	48	18	14	18	15	19	0	3
5	10	10	10	1	4	15	−1	1	20	14	13

Вычисляем

$$R = \sum_{i=1}^{19}\left(R_i - \frac{20+1}{2}\right)\cdot\left(R_{i+1} - \frac{20+1}{2}\right) =$$
$$= (12 - 10{,}5)\cdot(8 - 10{,}5)\cdots(3 - 10{,}5)\cdot(13 - 10{,}5) = 35{,}25;$$

$$\mathbf{D}(R) = \frac{20^2\cdot 21\cdot 17\cdot(5\cdot 20 + 6)}{720} = 21033{,}3 \quad (\sqrt{\mathbf{D}(R)} = 144{,}99);$$

$$|R^*| = \frac{35{,}25}{144{,}99} = 0{,}249 < u_{0,975} = 1{,}96,$$

и гипотеза случайности не отклоняется.

4.3.9.2. Критерий Бартелса

Пусть R_i — ранг i-го наблюдения в последовательности n наблюдений x_i. Бартелсом [516] рассмотрен ранговый критерий случайности ряда, основанный на статистике

$$B = \frac{\sum_{i=1}^{n-1} (R_i - R_{i+1})^2}{\sum_{i=1}^{n} (R_i - \bar{R})^2}.$$

При совпадении элементов выборки ранги следует случайным образом распределить среди них. Значение B заключено в интервале $\frac{12}{n(n+1)} < B < 4 - \frac{12}{n(n+1)}$.

В пределе при $n \to \infty$ имеем $0 < B < 4$. Моменты распределения B равны [516]: $\mathbf{M}(B) = 2$; $\mathbf{D}(B) = \frac{4(n-2)(5n^2-2n-9)}{5n(n+1)(n-1)^2}$. Коэффициент асимметрии распределения B равен $\alpha_3 = \frac{12}{35n\sqrt{n}}$, коэффициент эксцесса $\alpha_4 \approx 3 - \frac{798}{175n}$. Отсюда следует, что распределение симметрично, но медленно стремится к нормальному эксцессу.

Гипотеза случайности отклоняется на уровне значимости α, если $(2 - B_\alpha) < B < (B_\alpha + 2)$. Критические точки B_α определяются по формуле [516] $B_\alpha = a + bn^c(\ln n)^d$, где a, b, c, d — коэффициенты, приведенные в табл. 182.

Таблица 182

Коэффициенты аппроксимации критического значения критерия Бартелса

(α — уровень значимости) [516]

α	0,01	0,025	0,05	0,1
a	−0,023	−0,004	0,119	−0,465
b	0,261	0,381	0,440	1,184
c	−0,345	−0,266	−0,230	−0,088
d	2,212	1,748	1,520	0,674

При $n \to \infty$ $\mathbf{M}(B) = 2$ и $\mathbf{D}(B) = \frac{4}{n}$, или более точно $\mathbf{D}(B) = \frac{20}{5n+7}$.

При $n > 100$ критерий приобретает вид $B^* = \frac{B - \mathbf{M}(B)}{\sqrt{\mathbf{D}(B)}} = \frac{B-2}{2\sqrt{\frac{5}{5n+7}}}$.

Если $|B^*| < u_{\frac{1+\alpha}{2}}$, то гипотеза случайности не отклоняется.

Задача 288. *Проверить гипотезу случайности ряда значений, заданных в задаче 285, критерием Бартелса.*

Вычисляем (используем ранги, приведенные в задаче 287)

$$\bar{R} = \frac{1}{n} \cdot \sum_{i=1}^{n} R_i = \frac{n+1}{2} = 10{,}5; \quad \sum_{i=1}^{19} (R_i - R_{i+1})^2 = (12-8)^2 + \ldots + (3-13)^2 = 1327;$$

$$\sum_{i=1}^{19} \left(R_i - \bar{R}\right)^2 = (12-10{,}5)^2 + \ldots + (3-10{,}5)^2 = 614{,}5;$$

$$B = \frac{\sum_{i=1}^{19} (R_i - R_{i+1})^2}{\sum_{i=1}^{19} (R_i - \bar{R})^2} = \frac{1327}{614{,}5} = 2{,}159.$$

При $n = 20$ и $\alpha = 0{,}05$ из табл. 182 имеем $a = 0{,}119$; $b = 0{,}440$; $c = -0{,}230$ и $d = 1{,}520$. Тогда $B_{0,05} = 0{,}119 + 0{,}440 \cdot 20^{-0,230} \cdot (\ln 20)^{1,520} = 1{,}29$.

Так как $(2 - 1{,}29) = 0{,}71 < B = 2{,}159 < (2 + 1{,}29) = 3{,}29$, гипотеза случайности не отклоняется.

4.3.10. Критерий кумулятивной суммы

В [517, 518] изучен критерий отсутствия тренда, основанный на сумме

$$V = \sum_{i=1}^{n} \delta(x_i - \tilde{x}), \quad \text{где} \quad \tilde{x}\text{ — медиана и} \quad \delta(z) = \begin{cases} 1, & \text{если } z \geqslant 0; \\ -1, & \text{если } z < 0. \end{cases}$$

Статистикой критерия является R — число переходов через нуль суммы V. Критерий тренда отклоняется на уровне значимости α, если $R_1(\alpha) < R < R_2(\alpha)$, где $R_1(\alpha)$ и $R_2(\alpha)$ — критические значения, приведенные в табл. 183.

Таблица 183

Критические значения $R_1(\alpha)$ и $R_2(\alpha)$ критерия кумулятивной суммы [517]

n	Доверительная вероятность α				n	Доверительная вероятность α			
	0,90		0,95			0,90		0,95	
	R_1	R_2	R_1	R_2		R_1	R_2	R_1	R_2
10	0	6	0	6	200	4	33	2	36
20	0	9	0	10	300	4	40	3	44
30	1	12	0	12	400	5	47	4	52
40	1	14	0	15	500	6	53	4	58
50	1	15	1	17	600	7	58	5	64
60	1	17	1	18	700	7	63	5	69
70	2	19	1	20	800	8	67	5	74
80	2	20	1	22	900	9	71	6	79
90	2	21	1	23	1000	9	75	6	83
100	2	23	1	25					

Задача 289. *Проверить гипотезу случайности ряда значений x, заданных в задаче 285, критерием кумулятивной суммы.*

В нашем случае медиана ряда есть $\tilde{x} = 11$. Последовательность значений $\delta(x_i - \tilde{x})$ имеет вид:

$$+1,\ -1,\ -1,\ -1,\ -1,\ +1,\ +1,\ +1,\ +1,\ -1,\ -1,\ -1,\ -1,\ +1,\ -1,\ -1,\ +1,\ +1,\ -1,\ +1.$$

Этот ряд порождает последовательность сумм V_i:

$$+1,\ 0,\ -1,\ -2,\ -3,\ -2,\ -1,\ 0,\ +1,\ 0,\ -1,\ -2,\ -3,\ -2,\ -3,\ -4,\ -3,\ -2,\ -3,\ -2.$$

Следовательно, $R = 3$ (три нуля). Из табл. 183 для $n = 20$ и $\alpha = 0{,}95$ находим $R_1 = 0$ и $R_2 = 10$.

Так как $R_1 = 0 < R = 3 < R_2 = 10$, исследуемый ряд должен быть признан случайным.

4.3.11. Знаково-ранговый критерий Холлина

Холлин опубликовал в последнее время ряд работ [552–555], посвященных детальному изучению проблемы использования рангов в критериях случайности. Укажем на последний результат [556], в котором предложено в определенном смысле обобщение рангового критерия Вальда–Волфовитца (см. раздел 4.3.9.1) и знакового критерия кумулятивной суммы (см. раздел 4.3.10). Предложенный Холлином знаково-ранговый критерий автокорреляции основан на статистике

$$r = \frac{1}{k(n-1)} \sum_{i=2}^{n} \delta[(x_i - \tilde{x})(x_{i-1} - \tilde{x})] R_i R_{i-1},$$

где k — коэффициент, зависящий от объема выборки (некоторые его значения приведены в табл. 184); $\tilde{x}$ — медиана выборочного ряда $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$; R_i — ранг величины $z_i = |x_i - \tilde{x}|$ в общем упорядоченном по возрастанию ряду значений

$$z_1, z_2, z_n;\ \delta(y) = \begin{cases} 1, & \text{если } y > 0; \\ -1, & \text{если } y < 0; \\ 0, & \text{если } y = 0. \end{cases}$$

Таблица 184

Значения k для знаково-рангового критерия Холлина [556]

n	5	10	20	50	100	200	400
k	10,11	36,95	140,62	851,62	3370	13407	53480

Ряд значений x_i признается случайным, если $|r| < r_\alpha$, где r_α — критические значения, приведенные в табл. 185 (α — уровень значимости). Критерий Холлина обладает большей эффективностью среди всех, рассмотренных ранее ранговых критериев.

Задача 290. *Проверить гипотезу случайности ряда значений x_i, заданных в задаче 285, знаково-ранговым критерием Холлина на уровне значимости $\alpha = 0{,}05$.*

Медиана ряда равна $\tilde{x} = \dfrac{10+12}{2} = 11$. Напомним, что медиана ряда равна полусумме центральных значений упорядоченного по возрастанию ряда при n четном и центральному значению при n нечетном. Составляем ряд значений z_i (R_i — ранг z_i в последовательности) и соответствующих им значений $\delta_{1i} = \delta(x_i - \tilde{x})$ и $\delta_{2i} = \delta[(x_i - \tilde{x})\,(x_{i-1} - \tilde{x})] = \delta_{1i} \cdot \delta_{1,i-1}$:

z_i	R_i	δ_{1i}	δ_{2i}	z_i	R_i	δ_{1i}	δ_{2i}	z_i	R_i	δ_{1i}	δ_{2i}	z_i	R_i	δ_{1i}	δ_{2i}
2	3,5	1		6	8	1	−1	1	1,5	1	−1	14	16	−1	1
3	5,5	−1	−1	10	12,5	1	1	2	3,5	−1	−1	38	19	1	−1
4	7	−1	1	23	17	1	1	8	11	−1	1	39	20	1	1
7	9,5	−1	1	37	18	1	1	7	9,5	1	−1	11	14	−1	−1
1	1,5	−1	1	10	12,5	−1	−1	12	15	−1	−1	3	5,5	1	−1

Для равных значений z_i применяем средние ранги.

Из табл. 184 для $n = 20$ находим $k = 140{,}62$ и вычисляем статистику критерия

$$r = \frac{1}{140{,}62 \cdot (20-1)} \cdot \sum \delta[(x_i - \tilde{x}) \cdot (x_{i-1} - \tilde{x})] \cdot R_i \cdot R_{i-1} =$$
$$= 3{,}7428 \cdot 10^{-4} \cdot (-3{,}5 \cdot 5{,}5 + 5{,}5 \cdot 7 + 7 \cdot 9{,}5 + 9{,}5 \cdot 1{,}5 - \ldots$$
$$\ldots + 19 \cdot 20 - 20 \cdot 14 - 14 \cdot 5{,}5) = 0{,}078.$$

Из табл. 185 для $n = 20$ и $\alpha = 0{,}05$ находим критическое значение $r_{0,05} = 0{,}378$. Так как $r = 0{,}078 < r_{0,05} = 0{,}378$, ряд следует признать случайным.

Таблица 185

Критические значения r_α знаково-рангового критерия Холлина [556]

n	Уровень значимости α			n	Уровень значимости α		
	0,10	0,05	0,025		0,10	0,05	0,025
5	0,667	0,816	0,915	28	0,250	0,317	0,374
6	0,608	0,749	0,834	29	0,245	0,311	0,367
7	0,548	0,681	0,778	30	0,241	0,306	0,361
8	0,501	0,625	0,720	32	0,233	0,296	0,349
9	0,467	0,587	0,682	34	0,225	0,287	0,339
10	0,421	0,553	0,642	36	0,219	0,279	0,320
11	0,415	0,522	0,610	38	0,213	0,271	0,320
12	0,394	0,498	0,583	40	0,207	0,264	0,312
13	0,380	0,479	0,558	42	0,202	0,257	0,304
14	0,361	0,455	0,535	44	0,197	0,251	0,297
15	0,349	0,442	0,518	46	0,192	0,246	0,291
16	0,338	0,428	0,502	48	0,188	0,240	0,285
17	0,325	0,412	0,485	50	0,184	0,235	0,279
18	0,317	0,402	0,472	55	0,175	0,224	0,266
19	0,306	0,388	0,459	60	0,168	0,214	0,254
20	0,298	0,378	0,446	65	0,161	0,206	0,244
21	0,292	0,370	0,436	70	0,155	0,198	0,235
22	0,283	0,361	0,425	75	0,150	0,191	0,227
23	0,276	0,350	0,413	80	0,145	0,185	0,220
24	0,270	0,343	0,404	85	0,140	0,180	0,213
25	0,265	0,338	0,400	90	0,136	0,174	0,207
26	0,260	0,330	0,388	95	0,133	0,170	0,202
27	0,255	0,323	0,323	100	0,129	0,165	0,197

4.3.12. Критерии обнаружения выбросов

Среди методов оценки однородности выборочных данных следует выделить группу методов, связанных с обнаружением аномальных, не согласующихся с остальными элементами выборки наблюдений.

Результаты проведенных экспериментов или испытаний иногда существенно отличаются от наблюдаемых средних значений. Необходимо быть уверенным, что эти результаты не являются грубым промахом, ошибкой при фиксировании наблюдаемой величины. Другими словами, следует убедиться, являются ли эти отклонения случайными, либо их появление является следствием проявления систематических (или, по крайней мере, фиксированных) неслучайных процессов.

Для проверки значимости подозрительных экспериментальных данных разработаны специальные статистические критерии. Если такой критерий подтверждает гипотезу о том, что подозрительный результат значимо отличается от остальных, то исследователь должен выявить причину такого отклонения. Если причина содержится в нарушении условий эксперимента (скачок напряжения сети, поломка измерительного прибора), то она должна быть устранена, а полученный выброс исключается из выборки. Возможно, что причина отклонения содержится в появлении нового физического процесса. Тогда статистическое установление значимости выброса привлечет внимание исследователя к этому процессу.

Рекомендуется проводить обязательный анализ значимости отклонения крайних значений выборки от остальных, так как если они являются выбросами, то их использование при оценке выборочных моментов (особенно высших порядков) и проверка различных статистических гипотез может привести к большим ошибкам.

В рассматриваемых ниже критериях предполагается, что выборки предварительно ранжированы по возрастанию.

4.3.12.1. Критерии выбросов в случае нормального распределения

4.3.12.1.1. Критерий Шовене

Согласно критерию Шовене [519, 520] элемент выборки x_i объема n является выбросом, если вероятность его отклонения от среднего значения не больше $\frac{1}{12n}$. Критические значения K^* статистики Шовене

$$K = \frac{|x_i - \bar{x}|}{s}, \quad \text{где} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad s = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

приведены в табл. 186.

Таблица 186

Критические значения K^* критерия Шовене [519]

n	4	5	6	10	15	25	50	100	300
K^*	1,54	1,65	1,73	1,96	2,13	2,33	2,57	2,81	3,14

Если $K > K^*$, то значение x_i $(i = 1, \ldots, n)$ должно быть признано выбросом.

Задача 291. *Пусть в нашем распоряжении имеется выборка объема* $n = 10$:

$$x_i: \ -3,\ 2,\ 6,\ 11,\ 15,\ 19,\ 26,\ 28,\ 30,\ 41.$$

Необходимо проверить наличие выброса в выборке критерием Шовене.

Имеем $\bar{x} = 17{,}5$; $s^2 = 192{,}71$ $(s = 13{,}882)$. Вычисляем (для $x_i = 41$)

$$K = \frac{41 - 17{,}5}{13{,}882} = 1{,}693.$$

Для $n = 10$ из табл. 186 имеем $K^* = 1{,}96$.

Так как $K = 1{,}693 < K^* = 1{,}96$, крайнее значение в выборке не является выбросом. Очевидно, что проверка крайнего левого значения приведет к аналогичному результату.

4.3.12.1.2. Критерий Ирвина

Предложен в [521], используется в случае, когда дисперсия распределения известна заранее. Статистика критерия имеет вид

$$\tau = \frac{x_n - x_{n-1}}{\sigma}$$

— для проверки наибольшего значения x_n, и

$$\tau^* = \frac{x_2 - x_1}{\sigma}$$

— если подозрительным является наименьшее значение x_1. При τ $(\tau^*) < \tau(\alpha)$ — наибольшее (наименьшее) значение признается выбросом с вероятностью α. Критические значения $\tau(\alpha)$ приведены в табл. 187.

Таблица 187

Критические значения $\tau(\alpha)$ критерия Ирвина [29]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
2	2,33	2,77	3,64	70	0,84	1,05	1,53
3	1,79	2,17	2,90	80	0,83	1,04	1,50
10	1,18	1,46	2,03	90	0,82	1,03	1,50
20	1,03	1,27	1,80	100	0,81	1,02	1,47
30	0,96	1,20	1,70	200	0,75	0,95	1,38
40	0,91	1,15	1,63	300	0,72	0,91	1,32
50	0,88	1,11	1,60	500	0,68	0,87	1,28
60	0,86	1,08	1,57	1000	0,65	0,83	1,22

Задача 292. *Проверить гипотезу о наличии выбросов для данных задачи* 291 *при условии, что* $\sigma = 17$ (*принять* $\alpha = 0{,}95$).

Имеем $\tau = \dfrac{41-30}{17} = 0{,}647$ и $\tau^* = \dfrac{2+3}{17} = 0{,}294$.

Из табл. 187 для $n = 10$ и $\alpha = 0{,}95$ имеем $\tau(0{,}95) = 1{,}46$.

Так как $\tau = 0{,}647$ ($\tau^* = 0{,}294$) $< \tau(0{,}95) = 1{,}46$, крайние значения не являются выбросами.

4.3.12.1.3. Критерий Груббса

Груббсом в [522] предложена серия критериев для обнаружения выбросов, основанная на статистиках

$$\tau_1 = \frac{x_n - \bar{x}}{\sigma}; \quad \tau_1^* = \frac{x_n - \bar{x}}{s}$$

если подозрительным является наибольшее значение x_n, и

$$\tau_2 = \frac{\bar{x} - x_1}{\sigma}; \quad \tau_2^* = \frac{\bar{x} - x_1}{s}$$

если подозрительным является наименьшее значение x_1.

Статистики τ_1 или τ_2 применяются, когда дисперсия известна заранее; статистики τ_1^* и τ_2^* — когда дисперсия оценивается по выборке с помощью соотношения $s^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$.

При τ_1 (τ_1^*) $\geqslant \tau_1(\alpha)$ или τ_2 (τ_2^*) $\geqslant \tau_2(\alpha)$ проверяемое значение (x_n при τ_1 и τ_1^*, x_1 при τ_2 и τ_2^*) признается выбросом. Критические значения $\tau_1(\alpha)$ и $\tau_2(\alpha)$ приведены в табл. 188.

При $n > 25$ критические значения τ_1 можно найти, используя приближение [29]

$$\tau_1(\alpha) = u_{1+\frac{\alpha-1}{n}}\sqrt{\frac{n-1}{n}}.$$

где u_γ — γ-квантиль стандартного нормального распределения.

Для $\tau_2(\alpha)$ предложена аппроксимация [36] при $\alpha = 0{,}95$

$$\tau_2(0,95) = \begin{cases} 1{,}31 + 0{,}435\ln(n-2{,}7) & \text{при } 5 \leqslant n < 35; \\ 1{,}962 + 0{,}281\ln(n-15) & \text{при } 35 \leqslant n \leqslant 500. \end{cases}$$

Таблица 188

Критические значения $\tau_1(\alpha)$, $\tau_2(\alpha)$ и $\tau_3(\alpha)$ статистик Груббса [25, 29]

n	Доверительная вероятность α								
	0,90			0,95			0,99		
	τ_1	τ_2	τ_3	τ_1	τ_2	τ_3	τ_1	τ_2	τ_3
3	1,497	1,406	1,818	1,738	1,412	2,121	2,215	1,414	2,712
4	1,696	1,645	1,943	1,941	1,689	2,234	2,431	1,710	2,806
5	1,835	1,791	2,036	2,080	1,869	2,319	2,574	1,917	2,877
6	1,939	1,894	2,111	2,184	1,996	2,386	2,679	2,067	2,934
7	2,022	1,974	2,172	2,267	2,093	2,442	2,761	2,182	2,981
8	2,091	2,041	2,224	2,334	2,172	2,490	2,828	2,273	3,022
9	2,150	2,097	2,269	2,392	2,237	2,531	2,884	2,349	3,057
10	2,200	2,146	2,309	2,441	2,294	2,568	2,931	2,414	3,089
11	2,245	2,190	2,344	2,484	2,343	2,601	2,973	2,470	3,117
12	2,284	2,229	2,376	2,523	2,387	2,630	3,010	2,519	3,143
13	2,320	2,264	2,406	2,557	2,426	2,657	0,043	2,562	3,166
14	2,352	2,297	2,432	2,589	2,461	2,682	3,072	2,602	3,187
15	2,382	2,326	2,457	2,617	2,493	2,705	3,099	2,638	3,207
16	2,409	2,354	2,480	2,644	2,523	2,726	3,124	2,670	3,226
17	2,434	2,380	2,502	2,668	2,551	2,746	3,147	2,701	3,243
18	2,458	2,404	2,522	2,691	2,577	2,765	3,168	2,728	3,259
19	2,480	2,426	2,541	2,712	2,600	2,783	3,188	2,754	3,275
20	2,500	2,447	2,559	2,732	2,623	2,799	3,207	2,778	3,289
21	2,529	2,467	2,576	2,750	2,644	2,815	3,224	2,801	3,303
22	2,538	2,486	2,592	2,768	2,664	2,830	3,240	2,823	3,316
23	2,555	2,504	2,607	2,784	12,683	2,844	3,255	2,843	3,328
24	2,571	2,520	2,621	2,800	2,701	2,857	3,269	2,862	3,340
25	2,587	2,537	2,635	2,815	2,717	2,870	3,282	2,880	3,351

Если наряду с дисперсией заранее известно и среднее значение (μ) совокупности, из которой извлекается выборка наряду с дисперсией, то используется статистика

$$\tau_3 = \frac{|x_n - \mu|}{\sigma},$$

критические значения которой $\tau_3(\alpha)$ также приведены в табл. 188. Если $\tau_3 \geqslant \tau_3(\alpha)$, то выброс признается значимым.

Задача 293. *Проверить гипотезу о наличии выбросов в выборке задачи* 291 *критериями Груббса при* $\alpha = 0{,}95$ (*для критериев* τ_1, τ_2 *и* τ_3 *будем считать, что* $\sigma = 14$, *для критерия* τ_3 *принять* $\mu = 17$).

Вычисляем

$$s = 13{,}697; \quad \bar{x} = 17{,}4; \quad \tau_1 = \frac{41 - 17{,}4}{14} = 1{,}686; \quad \tau_1^* = \frac{41 - 17{,}4}{13{,}697} = 1{,}723;$$

$$\tau_2 = \frac{17{,}4 + 3}{14} = 1{,}457; \quad \tau_2^* = \frac{17{,}4 + 3}{13{,}697} = 1{,}489; \quad \tau_3 = \frac{41 - 17}{14} = 1{,}714,$$

Из табл. 188 для $\alpha = 0{,}95$ и $m = 10$ находим

$$\tau_1(0{,}95) = 2{,}441, \quad \tau_2(0{,}95) = 2{,}294 \quad \text{и} \quad \tau_3(0{,}95) = 2{,}568.$$

Так как $\tau_1 = 1{,}686$ ($\tau_2 = 1{,}457$) $< \tau_1(0{,}95) = 2{,}441$; $\tau_1^* = 1{,}723$ ($\tau_2^* = 1{,}489$) $< \tau_2(0{,}95) = 2{,}294$ и $\tau_3 = 1{,}714 < \tau_3(0,95) = 2,568$, наличие выбросов в выборке отклоняется.

Воспользуемся аппроксимацией для $\tau_2(0{,}95)$:

$$\tau_2(0{,}95) = 1{,}31 + 0{,}435 \cdot \ln(10 - 2{,}7) = 2{,}175,$$

что близко к табличному значению $\tau_2(0{,}95) = 2{,}294$.

4.3.12.1.4. Критерий наибольшего абсолютного отклонения

Основан на статистике

$$\tau_4 = \frac{\max\limits_{1 \leqslant i \leqslant n} |x_i - \bar{x}|}{s}, \quad \text{где} \quad s = \left\{ \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}^{\frac{1}{2}}.$$

При $\tau_4 \geqslant \tau_4(\alpha)$ значение x_i признается выбросом (критические значения $\tau_4(\alpha)$ приведены в табл. 189).

Таблица 189

Критические значения $\tau_4(\alpha)$ критерия наибольшего абсолютного отклонения [25]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
3	1,412	1,414	1,414	14	2,461	2,602	2,859
4	1,689	1,710	1,728	16	2,523	2,670	2,946
5	1,869	1,917	1,972	18	2,577	2,718	3,017
6	1,996	2,067	2,161	20	2,623	2,779	3,079
7	2,093	2,182	2,310	22	2,664	2,823	3,132
8	2,172	2,273	2,431	24	2,701	2,862	3,179
9	2,238	2,349	2,532	25	2,734	2,897	3,220
10	2,294	2,414	2,616	28	2,764	2,929	3,258
12	2,387	2,519	2,753	30	2,792	2,958	3,291

В [36] предложена весьма точная аппроксимация (с точностью до $\leqslant 0{,}1$) для $\tau_4(\alpha)$ при $\alpha = 0{,}95$:

$$\tau_4(0{,}95) = \begin{cases} 1{,}39 + 0{,}462 \ln(n-3) & \text{при } 5 \leqslant n < 35; \\ 2{,}136 - 0{,}281 \ln(n-15) & \text{при } 35 \leqslant n \leqslant 500. \end{cases}$$

Задача 294. *Проверить гипотезу наличия выбросов в выборке задачи 291 критерием наибольшего абсолютного отклонения при доверительной вероятности $\alpha = 0{,}95$.*

Вычисляем $\tau_4 = \dfrac{\max|x_i - \bar{x}|}{s} = \dfrac{41 - 17{,}5}{13{,}679} = 1{,}718$.

Из табл. 189 для $n = 10$ и $\alpha = 0{,}95$ имеем $\tau_4(0{,}95) = 2{,}414$.

Так как $\tau_4 = 1{,}718 < \tau_4(0{,}95) = 2{,}414$, гипотеза о наличии выбросов не находит подтверждения.

Вычислим теперь аппроксимацию

$$\tau_4(0{,}95) = 1{,}39 + 0{,}462 \cdot \ln(10-3) = 2{,}289,$$

что близко к табличному значению.

4.3.12.1.5. Критерий Дэвида

Является модификацией критерия Груббса (см. раздел 4.3.12.1.3), использует статистику [523] $\tau_5 = \dfrac{x_n - \bar{x}}{s}$ или $\tau_5^* = \dfrac{\bar{x} - x_1}{s}$, где $s^2 = \dfrac{1}{m-1} \sum\limits_{i=1}^{m} (y_i - \bar{y})^2$ — выборочная дисперсия, оцениваемая по отдельной независимой выборке (объема m) значений y_i.

Критические значения $\tau_5(\alpha)$ статистики Дэвида приведены в табл. 190. При $n > 25$ применима аппроксимация

$$\tau_5(\alpha) \approx u_{1+\frac{\alpha-1}{n}} \left(1 + \frac{3}{m-1}\sqrt{\frac{n-1}{n}}\right),$$

где u_γ — γ-квантиль стандартного нормального распределения.

При τ_5 $(\tau_5^*) \geqslant \tau_5(\alpha)$ одно из крайних выборочных значений признается выбросом с вероятностью α.

Задача 295. *Для ряда данных задачи 291 проверить наличие выбросов критерием Дэвида при доверительной вероятности $\alpha = 0{,}95$. Известно, что оценка дисперсии $s^2 = 182{,}15$ получена по независимой выборке объема $m = 10$.*

Имеем $s = 13{,}496$, тогда $\tau_5 = \dfrac{41 - 17{,}5}{13{,}496} = 1{,}741$ и $\tau_5^* = \dfrac{17{,}5 + 3}{13{,}496} = 1{,}519$.

Из табл. 190 для $n = 10$, $m = 10$ и $\alpha = 0{,}95$ находим $\tau_5(0{,}95) = 2{,}89$.

Так как $\tau_5 = 1{,}74 < \tau_5(0{,}95) = 2{,}89$, гипотеза о наличии выбросов в выборке отклоняется.

Такой же результат дает аппроксимация. Имеем

$$u_{1+\frac{\alpha-1}{n}} = u_{1+\frac{0{,}95-1}{10}} = u_{0{,}995} \approx 4{,}91 \cdot \left(0{,}995^{0{,}14} - 0{,}005^{0{,}14}\right) = 2{,}568;$$

$$\tau_5(0{,}95) \approx 2{,}568 \cdot \left(1 + \frac{3}{9} \cdot \sqrt{\frac{9}{10}}\right) = 3{,}38.$$

Видим, что ошибка аппроксимации велика (точное значение $\tau_5(0{,}95) = 2{,}89$). Это связано с малым объемом выборки $n = 10$ (для точной аппроксимации необходимо $n > 25$).

4.3.12.1.6. Критерии Диксона

Используются для быстрого выявления выпадающих наблюдений по отношению размаха и подразмахов.

Статистиками критериев являются [524] (в скобках указаны проверяемые наблюдения):

— для проверки одного сомнительного наблюдения

$$r_{10} = \frac{x_2 - x_1}{x_n - x_1} \quad (\text{для проверки } x_1); \qquad r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1} \quad (\text{для проверки } x_n);$$

— для проверки одного сомнительного наблюдения независимо от противоположного крайнего наблюдения

$$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1} \quad (x_1); \qquad r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2} \quad (x_n);$$

— для проверки одного сомнительного наблюдения независимо от двух противоположных крайних

$$r_{12} = \frac{x_2 - x_1}{x_{n-2} - x_1} \quad (x_1); \qquad r_{12} = \frac{x_n - x_{n-1}}{x_n - x_3} \quad (x_n);$$

— для проверки одного сомнительного наблюдения независимо от следующего по величине

$$r_{20} = \frac{x_3 - x_1}{x_n - x_1} \quad (x_1); \qquad r_{20} = \frac{x_n - x_{n-2}}{x_n - x_1} \quad (x_n);$$

— для проверки одного сомнительного наблюдения независимо от следующего по величине и крайнего противоположного

$$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1} \quad (x_1); \qquad r_{21} = \frac{x_n - x_{n-2}}{x_n - x_2} \quad (x_n);$$

Таблица 190

Критические значения $\tau_5(\alpha)$ статистики Дэвида

n	m								
	4	5	6	7	8	9	10	11	13
Доверительная вероятность $\alpha = 0{,}90$									
10	1,68	1,92	2,09	2,23	2,33	2,42	2,50	2,56	2,68
11	1,66	1,90	2,07	2,20	2,30	2,39	2,46	2,53	2,64
12	1,65	1,88	2,05	2,17	2,28	2,36	2,44	2,50	2,61
13	1,63	1,86	2,03	2,16	2,26	2,34	2,41	2,47	2,58
14	1,62	1,85	2,01	2,14	2,24	2,32	2,39	2,45	2,56
15	1,61	1,84	2,00	2,12	2,22	2,31	2,38	2,44	2,54
16	1,61	1,83	1,99	2,11	2,21	2,29	2,36	2,42	2,52
17	1,60	1,82	1,98	2,10	2,20	2,28	2,35	2,41	2,51
18	1,59	1,82	1,97	2,09	2,19	2,26	2,34	2,39	2,49
19	1,59	1,81	1,96	2,08	2,18	2,26	2,33	2,38	2,48
20	1,58	1,80	1,96	2,08	2,17	2,25	2,32	2,37	2,47
24	1,57	1,78	1,94	2,05	2,15	2,22	2,29	2,34	2,44
30	1,55	1,77	1,92	2,03	2,12	2,20	2,26	2,32	2,41
40	1,54	1,75	1,90	2,01	2,10	2,17	2,23	2,29	2,38
60	1,52	1,73	1,87	1,98	2,07	2,14	2,20	2,26	2,35
120	1,51	1,71	1,85	1,96	2,05	2,12	2,18	2,23	2,32
∞	1,50	1,70	1,83	1,94	2,02	2,09	2,15	2,20	2,28
Доверительная вероятность $\alpha = 0{,}95$									
10	2,01	2,27	2,46	2,60	2,72	2,81	2,89	2,96	3,08
11	1,98	2,24	2,42	2,56	2,67	2,76	2,84	2,91	3,03
12	1,96	2,21	2,39	2,52	2,63	2,72	2,80	2,87	2,98
13	1,94	2,19	2,35	2,50	2,60	2,69	2,76	2,83	2,94
14	1,93	2,17	2,34	2,47	2,57	2,66	2,74	2,80	2,91
15	1,91	2,15	2,32	2,45	2,55	2,64	2,71	2,77	2,88
16	1,90	2,14	2,31	2,43	2,53	2,62	2,69	2,75	2,86
17	1,89	2,13	2,29	2,42	2,52	2,60	2,67	2,73	2,84
18	1,88	2,11	2,28	2,40	2,50	2,58	2,65	2,71	2,82
19	1,87	2,11	2,27	2,39	2,49	2,57	2,64	2,70	2,80
20	1,87	2,10	2,26	2,38	2,47	2,56	2,63	2,68	2,78
24	1,84	2,07	2,23	2,34	2,44	2,52	2,58	2,64	2,74
30	1,82	2,04	2,20	2,31	2,40	2,48	2,54	2,60	2,69
40	1,80	2,02	2,17	2,28	2,37	2,44	2,50	2,56	2,65
60	1,78	1,99	2,14	2,25	2,33	2,41	2,47	2,52	2,61
120	1,76	1,96	2,11	2,22	2,30	2,37	2,43	2,48	2,57
∞	1,74	1,94	2,08	2,18	2,27	2,33	2,39	2,44	2,52
Доверительная вероятность $\alpha = 0{,}99$									
10	2,78	3,10	3,32	3,48	3,62	3,73	3,82	3,90	4,04
11	2,72	3,02	3,24	3,39	3,52	3,63	3,72	3,79	3,93
12	2,67	2,96	3,17	3,32	3,45	3,55	3,64	3,71	3,84
13	2,63	2,92	3,12	3,27	3,38	3,48	3,57	3,64	3,76
14	2,60	2,88	3,07	3,22	3,33	3,43	3,51	3,58	3,70
15	2,57	2,84	3,03	3,17	3,29	3,38	3,46	3,53	3,65
16	2,54	2,81	3,00	3,14	3,25	3,34	3,42	3,49	3,60
17	2,52	2,79	2,97	3,11	3,22	3,31	3,38	3,45	3,56
18	2,50	2,77	2,95	3,08	3,19	3,28	3,35	3,42	3,53
19	2,49	2,75	2,93	3,06	3,16	3,25	3,33	3,39	3,50
20	2,47	2,73	2,91	3,04	3,14	3,23	3,30	3,37	3,47
24	2,42	2,68	2,84	2,97	3,07	3,16	3,23	3,29	3,38
30	2,38	2,62	2,79	2,91	3,01	3,08	3,15	3,21	3,30
40	2,34	2,57	2,73	2,85	2,94	3,02	3,08	3,13	3,22
60	2,29	2,52	2,68	2,79	2,88	2,95	3,01	3,06	3,15
120	2,25	2,48	2,62	2,73	2,82	2,89	2,95	3,00	3,08
∞	2,22	2,43	2,57	2,68	2,76	2,83	2,88	2,93	3,01

— для проверки одного сомнительного наблюдения независимо от следующего по величине и двух крайних противоположных

$$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1} \quad (x_1); \qquad r_{22} = \frac{x_n - x_{n-2}}{x_n - x_3} \quad (x_n).$$

Критические значения статистик $r_{10}(\alpha)$, $r_{11}(\alpha)$, $r_{12}(\alpha)$, $r_{20}(\alpha)$, $r_{21}(\alpha)$ и $r_{22}(\alpha)$, превышение которых приводит к признанию наличия выбросов в выборке, приведены в табл. 191. Критерий r_{10} рекомендуется применять при $3 \leqslant n \leqslant 7$; критерий r_{11} — при $8 \leqslant n \leqslant 10$, критерий r_{21} — при $11 \leqslant n \leqslant 13$; критерий r_{22} — при $14 \leqslant n \leqslant 25$.

Задача 296. *Для выборочного ряда задачи 291 проверить гипотезу о наличии выбросов критериями Диксона при доверительной вероятности $\alpha = 0{,}95$.*

Вычисляем:

$$r_{10} = \frac{x_2 - x_1}{x_n - x_1} = \frac{2+3}{41+32} = 0{,}1136; \qquad r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1} = \frac{41-30}{41+3} = 0{,}25;$$

$$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1} = \frac{2+3}{30+3} = 0{,}1515; \quad r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2} = \frac{41-30}{41-2} = 0{,}282;$$

$$r_{12} = \frac{x_2 - x_1}{x_{n-2} - x_1} = \frac{2+3}{28+3} = 0{,}1513; \quad r_{12} = \frac{x_n - x_{n-1}}{x_n - x_3} = \frac{41-30}{41-6} = 0{,}3143;$$

$$r_{20} = \frac{x_3 - x_1}{x_n - x_1} = \frac{6+3}{41+3} = 0{,}2045; \qquad r_{20} = \frac{x_n - x_{n-2}}{x_n - x_1} = \frac{41-28}{41+3} = 0{,}2954;$$

$$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1} = \frac{6+3}{30+3} = 0{,}2727; \quad r_{21} = \frac{x_n - x_{n-2}}{x_n - x_2} = \frac{41-28}{41-2} = 0{,}3333;$$

$$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1} = \frac{6+3}{28+3} = 0{,}2903; \quad r_{22} = \frac{x_n - x_{n-2}}{x_n - x_3} = \frac{41-28}{41-6} = 0{,}3714.$$

Из табл. 191 для $n = 10$ и $\alpha = 0{,}95$ находим критические значения

$$r_{10}(0{,}95) = 0{,}412; \quad r_{20}(0{,}95) = 0{,}531; \quad r_{11}(0{,}95) = 0{,}477;$$

$$r_{21}(0{,}95) = 0{,}612; \quad r_{12}(0{,}95) = 0{,}537; \quad r_{22}(0{,}95) = 0{,}682.$$

Сравнением убеждаемся, что (рассматриваются максимальные величины)

$$r_{10} = 0{,}25 < r_{10}(0{,}95) = 0{,}412; \qquad r_{11} = 0{,}28 < r_{11}(0{,}95) = 0{,}477;$$

$$r_{12} = 0{,}314 < r_{12}(0{,}95) = 0{,}537; \quad r_{20} = 0{,}2954 < r_{20}(0{,}95) = 0{,}531;$$

$$r_{21} = 0{,}33 < r_{21}(0{,}95) = 0{,}612; \qquad r_{22} = 0{,}3714 < r_{22}(0{,}95) = 0{,}682.$$

Следовательно, ни одна из статистик не приводит к принятию гипотезы о наличии выбросов в выборке.

4.3.12.1.7. Критерий Хоглина–Иглевича

В работе авторов критерия [525] рассматривается правило выделения выпадающих наблюдений в нормально распределенных выборках с помощью порядковых статистик. Правило обнаружения выбросов формулируется следующим образом. Наблюдение признается выбросом, если его значение находится вне интервала, ограниченного величинами

$$(1+k)x_{[l]} - kx_{[n+1-l]} \quad \text{и} \quad (1+k)x_{[n+1-l]} - kx_{[l]},$$

где $x_{[i]}$ — i-я порядковая статистика (т. е. i-й по величине член выборки, упорядоченной по возрастанию).

Таблица 191

Критические значения $r_{10}(\alpha)$, $r_{11}(\alpha)$, $r_{12}(\alpha)$, $r_{20}(\alpha)$, $r_{21}(\alpha)$ и $r_{22}(\alpha)$ статистик Диксона (значения приведены последовательно по строкам в соответствии с порядком записи, т. е. верхняя строка — r_{10}, …, нижняя — r_{22}) [25]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
6	0,482	0,560	0,698	13	0,515	0,570	0,670
	0,609	0,689	0,805	14	0,294	0,349	0,450
	0,745	0,824	0,925		0,336	0,395	0,502
	0,670	0,736	0,836		0,369	0,432	0,542
	0,821	0,872	0,951		0,395	0,445	0,538
	0,965	0,983	0,995		0,448	0,501	0,593
7	0,434	0,507	0,637		0,492	0,546	0,641
	0,530	0,610	0,740	15	0,285	0,338	0,438
	0,636	0,712	0,836		0,323	0,381	0,486
	0,596	0,661	0,778		0,354	0,416	0,523
	0,725	0,780	0,885		0,382	0,430	0,522
	0,850	0,881	0,945		0,431	0,483	0,574
8	0,399	0,468	0,590		0,454	0,507	0,595
	0,479	0,554	0,683	16	0,277	0,320	0,426
	0,557	0,632	0,760		0,313	0,359	0,472
	0,545	0,607	0,710		0,341	0,388	0,508
	0,650	0,710	0,829		0,370	0,406	0,508
	0,745	0,803	0,890		0,416	0,433	0,557
9	0,370	0,437	0,555		0,454	0,490	0,595
	0,441	0,512	0,635	17	0,269	0,313	0,416
	0,504	0,580	0,701		0,303	0,349	0,460
	0,505	0,565	0,667		0,330	0,377	0,493
	0,594	0,657	0,776		0,359	0,397	0,495
	0,676	0,737	0,840		0,403	0,440	0,542
10	0,349	0,412	0,527		0,438	0,475	0,577
	0,409	0,477	0,597	18	0,263	0,306	0,407
	0,454	0,537	0,655		0,295	0,341	0,449
	0,474	0,531	0,632		0,320	0,367	0,480
	0,551	0,612	0,726		0,350	0,379	0,484
	0,620	0,682	0,791		0,391	0,428	0,529
11	0,332	0,392	0,502		0,424	0,462	0,561
	0,385	0,450	0,566	19	0,258	0,306	0,398
	0,431	0,502	0,619		0,288	0,341	0,439
	0,449	0,504	0,603		0,311	0,367	0,469
	0,517	0,576	0,679		0,341	0,379	0,473
	0,578	0,637	0,745		0,380	0,428	0,517
12	0,318	0,376	0,482		0,412	0,462	0,547
	0,367	0,428	0,541	20	0,252	0,300	0,391
	0,406	0,473	0,590		0,282	0,334	0,430
	0,429	0,481	0,579		0,303	0,358	0,458
	0,490	0,546	0,642		0,333	0,372	0,464
	0,543	0,600	0,704		0,371	0,419	0,506
13	0,305	0,361	0,465		0,401	0,450	0,535
	0,350	0,410	0,520	21	0,247	0,295	0,384
	0,387	0,451	0,554		0,276	0,327	0,421
	0,411	0,461	0,557		0,296	0,349	0,449
	0,467	0,521	0,615		0,326	0,365	0,455

Окончание таблицы 191

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
21	0,363	0,410	0,496	26	0,268	0,319	0,411
	0,391	0,440	0,524		0,300	0,338	0,422
22	0,242	0,290	0,378		0,331	0,376	0,457
	0,270	0,320	0,414		0,354	0,399	0,486
	0,290	0,342	0,440	27	0,224	0,269	0,353
	0,320	0,358	0,447		0,246	0,295	0,383
	0,356	0,402	0,487		0,263	0,314	0,405
	0,382	0,430	0,514		0,296	0,334	0,417
23	0,238	0,285	0,372		0,325	0,370	0,450
	0,265	0,314	0,407		0,348	0,393	0,475
	0,284	0,336	0,432	28	0,220	0,266	0,349
	0,314	0,352	0,440		0,243	0,291	0,378
	0,349	0,395	0,479		0,259	0,309	0,399
	0,374	0,421	0,505		0,292	0,330	0,412
24	0,234	0,281	0,367		0,320	0,365	0,444
	0,260	0,309	0,400		0,342	0,387	0,469
	0,278	0,330	0,423	29	0,218	0,263	0,345
	0,309	0,347	0,434		0,239	0,287	0,374
	0,343	0,388	0,471		0,255	0,305	0,394
	0,367	0,413	0,497		0,288	0,326	0,407
25	0,230	0,277	0,362		0,316	0,360	0,438
	0,255	0,304	0,394		0,337	0,381	0,463
	0,273	0,324	0,417	30	0,215	0,260	0,341
	0,304	0,343	0,428		0,236	0,283	0,369
	0,337	0,382	0,464		0,251	0,301	0,389
	0,360	0,406	0,489		0,285	0,322	0,402
26	0,227	0,273	0,357		0,312	0,355	0,433
	0,250	0,299	0,389		0,332	0,376	0,457

Для выбора значения l используются варианты

$$l_1 = \frac{1}{2}\left[\frac{n+3}{2}\right]; \quad l_2 = \left[\frac{n}{4}+\frac{5}{12}\right]; \quad l_3 = \left[\frac{n}{4}+\frac{1}{4}\right],$$

где $[\ldots]$ — целая часть числа.

Значения коэффициентов k для различных вариантов выбора значений l и вероятности отсутствия выброса α приведены в табл. 192.

Задача 297. *Проверить гипотезу о наличии выбросов в выборке задачи* 291 *критерием Хоглина–Иглевича при доверительной вероятности* $\alpha = 0{,}95$.

Имеем $l_1 = \frac{1}{2}\cdot\left[\frac{10+3}{2}\right] = 3$; $l_2 = \left[\frac{10}{4}+\frac{5}{12}\right] = 2$; $l_3 = \left[\frac{10}{4}+\frac{1}{2}\right] = 3$.

Для $\alpha = 0{,}95$ и $n = 10$ из табл. 192 находим $k = 2{,}4$ (для l_1); $k = 2{,}2$ (для l_2) и $k = 1{,}8$ (для l_3). Окончательно имеем

$$x_{[l_1]} - k\cdot\left(x_{[n+1-l_1]} - x_{[l_1]}\right) = x_3 - 2{,}4\cdot(x_8 - x_3) = 6 - 2{,}4\cdot(28-6) = -46{,}8;$$

$$x_{[n+1-l_1]} + 2{,}4\cdot\left(x_{[n+1-l_1]} - x_{[l_1]}\right) = x_8 + 2{,}4\cdot(x_8 - x_3) = 6 + 2{,}4\cdot(28-6) = 80{,}8.$$

Видим, что вне интервала $(-46{,}8; 80{,}8)$ нет элементов выборки. Очевидно, что при $l_2 = 2$ интервал становится еще шире: $(-59{,}6; 91{,}6)$.

Следовательно, с вероятностью 0,95 можно утверждать отсутствие выбросов в выборке.

Таблица 192

Значения коэффициента k критерия Хоглина–Иглевича [525]

n	Доверительная вероятность α						n	Доверительная вероятность α					
	0,90			0,95				0,90			0,95		
	l_1	l_2	l_3	l_1	l_2	l_3		l_1	l_2	l_3	l_1	l_2	l_3
7	2,3	1,7	1,5	3,0	2,3	2,0	17	2,2	1,8	1,7	2,6	2,1	2,0
8	1,8	1,6	1,4	2,2	2,1	1,8	18	2,0	1,9	1,7	2,3	2,1	2,4
9	2,7	1,7	1,4	3,3	2,1	1,8	19	2,2	1,9	1,8	2,6	2,3	2,2
10	2,0	1,8	1,5	2,4	0,2	1,8	20	1,9	1,8	2,2	2,3	2,1	2,1
11	2,2	1,8	1,7	2,7	2,2	2,1	30	2,0	1,9	1,9	2,2	2,2	2,1
12	1,8	1,8	1,6	2,2	2,1	2,0	40	2,0	2,0	1,9	2,2	2,2	2,2
13	2,3	1,8	1,6	2,8	2,2	1,9	50	2,0	2,0	1,9	2,2	2,2	2,2
14	2,0	1,9	1,7	2,3	2,2	2,0	75	2,1	2,0	2,0	2,3	2,2	2,2
15	2,1	1,8	1,7	2,5	2,2	2,1	100	2,1	2,0	2,0	2,2	2,2	2,2
16	1,9	1,9	1,7	2,3	2,2	2,1	200	2,2	2,2	2,2	2,4	2,4	2,4

4.3.12.1.8. Критерий Титьена–Мура для обнаружения нескольких выбросов

Критерий, предложенный Титьеном и Муром [526], является обобщением критерия Груббса (см. раздел 4.3.12.1.3) на случай выявления нескольких выбросов в выборке. Для выделения k наибольших выбросов используется статистика

$$L_k = \frac{\sum\limits_{i=1}^{n-k} (x_i - \bar{x}_k)^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}, \quad \text{где} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{и} \quad \bar{x}_k = \frac{1}{n-k}\sum_{i=1}^{n-k} x_i.$$

Для выделения k наименьших наблюдений используется статистика

$$L_k^* = \frac{\sum\limits_{i=k+1}^{n} (x_i - \bar{x}*_k)^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}, \quad \text{где} \quad \bar{x}_k^* = \frac{1}{n-k}\sum_{i=k+1}^{n} x_i.$$

Если подозрительными являются как наименьшие, так и наибольшие наблюдения, то для их обнаружения используется следующее правило. Находятся абсолютные отклонения $d_i = |x_i - \bar{x}|$ и ранжируются по возрастанию от d_1 до d_n. Обозначим через z_i выборочное значение x_i, для которого d_i является i-м по величине. Для проверки гипотезы исключения k наибольших по модулю наблюдений используется статистика

$$E_k = \frac{\sum\limits_{i=1}^{n-k} (z_i - \bar{z}_k)^2}{\sum\limits_{i=1}^{n} (z_i - \bar{z})^2}, \quad \text{где} \quad \bar{z}_k = \frac{1}{n-k}\sum_{i=1}^{n-k} z_i.$$

При использовании критерия Титьена–Мура рекомендуется сначала использовать статистику L_k (L_k^*), а затем E_k. Наличие выбросов признается значимым с достоверностью α, если

$$L_k \ (L_k^*) \leqslant L_k(\alpha) \quad \text{или} \quad E_k \leqslant E_k(\alpha),$$

где $L_k(\alpha)$ и $E_k(\alpha)$ — критические значения, приведенные в табл. 193 и 194.

Таблица 193

Критические значения $L_k(\alpha)$ и $L_k^*(\alpha)$ критерия Титьена–Мура [526, 527]

n	k									
	1	2	3	4	5	6	7	8	9	10
Доверительная вероятность $\alpha = 90$										
3	0,011									
4	0,098	0,003								
5	0,200	0,038								
6	0,280	0,091	0,020							
7	0,348	0,148	0,056							
8	0,404	0,200	0,095	0,038						
9	0,448	0,248	0,134	0,068						
10	0,490	0,287	0,170	0,098	0,051					
11	0,526	0,326	0,208	0,128	0,074					
12	0,555	0,361	0,240	0,159	0,103	0,062				
13	0,578	0,388	0,270	0,186	0,126	0,082				
14	0,600	0,416	0,298	0,212	0,150	0,104	0,068			
15	0,611	0,436	0,322	0,236	0,172	0,124	0,086			
16	0,631	0,458	0,342	0,260	0,194	0,144	0,104	0,073		
17	0,648	0,478	0,364	0,282	0,216	0,165	0,125	0,092		
18	0,661	0,496	0,384	0,302	0,236	0,184	0,142	0,108	0,080	
19	0,676	0,510	0,398	0,316	0,251	0,199	0,158	0,124	0,094	
20	0,688	0,530	0,420	0,339	0,273	0,220	0,176	0,140	0,110	0,085
25	0,732	0,588	0,489	0,412	0,350	0,296	0,251	0,213	0,180	0,152
30	0,766	0,637	0,523	0,472	0,411	0,339	0,316	0,276	0,240	0,210
35	0,792	0,673	0,586	0,516	0,458	0,410	0,365	0,328	0,294	0,262
40	0,812	0,702	0,622	0,554	0,499	0,451	0,408	0,372	0,338	0,307
45	0,826	0,724	0,648	0,586	0,533	0,488	0,447	0,410	0,378	0,348
50	0,840	0,744	0,673	0,614	0,562	0,518	0,477	0,442	0,410	0,380
Доверительная вероятность $\alpha = 95$										
3	0,003									
4	0,051	0,001								
5	0,125	0,018								
6	0,203	0,055	0,010							
7	0,273	0,106	0,032							
8	0,326	0,146	0,064	0,022						
9	0,372	0,194	0,099	0,045						
10	0,418	0,233	0,129	0,070	0,034					
11	0,454	0,270	0,162	0,098	0,054					
12	0,489	0,305	0,196	0,125	0,076	0,042				
13	0,517	0,337	0,224	0,150	0,097	0,060				
14	0,540	0,363	0,250	0,174	0,122	0,079	0,050			
15	0,556	0,387	0,276	0,197	0,140	0,097	0,066			
16	0,579	0,410	0,300	0,219	0,159	0,115	0,082	0,055		
17	0,594	0,427	0,322	0,240	0,181	0,136	0,100	0,072		
18	0,608	0,447	0,337	0,259	0,200	0,154	0,116	0,086	0,062	
19	0,624	0,462	0,354	0,277	0,209	0,168	0,130	0,099	0,074	
20	0,639	0,484	0,377	0,299	0,238	0,188	0,150	0,115	0,088	0,066
25	0,696	0,550	0,450	0,374	0,312	0,262	0,222	0,184	0,154	0,126
30	0,730	0,599	0,506	0,434	0,376	0,327	0,283	0,245	0,212	0,183
35	0,762	0,642	0,554	0,482	0,424	0,376	0,334	0,297	0,264	0,235
40	0,784	0,672	0,588	0,523	0,468	0,421	0,378	0,342	0,310	0,280
45	0,802	0,696	0,618	0,556	0,502	0,456	0,417	0,382	0,350	0,320
50	0,820	0,722	0,646	0,588	0,535	0,490	0,450	0,414	0,383	0,356

Таблица 194

Критические значения $E_k(\alpha)$ критерия Титьена–Мура [526, 527]

n	k									
	1	2	3	4	5	6	7	8	9	10
Доверительная вероятность $\alpha = 90$										
3	0,003									
4	0,005	0,002								
5	0,127	0,022								
6	0,204	0,056	0,009							
7	0,268	0,094	0,027							
8	0,328	0,137	0,053	0,016						
9	0,377	0,175	0,080	0,032						
10	0,420	0,214	0,108	0,052	0,022					
11	0,449	0,250	0,138	0,073	0,036					
12	0,485	0,278	0,162	0,094	0,052	0,026				
13	0,510	0,309	0,189	0,116	0,068	0,038				
14	0,538	0,337	0,216	0,138	0,085	0,052	0,029			
15	0,558	0,360	0,240	0,160	0,105	0,067	0,040			
16	0,578	0,384	0,263	0,182	0,122	0,082	0,053	0,032		
17	0,594	0,406	0,284	0,198	0,140	0,095	0,064	0,042		
18	0,610	0,424	0,304	0,217	0,156	0,110	0,076	0,051	0,034	
19	0,629	0,442	0,322	0,234	0,172	0,124	0,089	0,062	0,042	
20	0,644	0,460	0,336	0,252	0,188	0,138	0,102	0,072	0,051	0,035
25	0,693	0,528	0,417	0,331	0,264	0,210	0,168	0,132	0,103	0,080
30	0,730	0,582	0,475	0,391	0,325	0,270	0,224	0,186	0,154	0,126
35	0,763	0,624	0,523	0,443	0,379	0,324	0,276	0,236	0,202	0,172
40	0,784	0,657	0,562	0,486	0,422	0,367	0,320	0,278	0,243	0,212
45	0,803	0,684	0,593	0,522	0,459	0,406	0,360	0,320	0,284	0,252
50	0,820	0,708	0,622	0,552	0,492	0,440	0,396	0,355	0,319	0,287
Доверительная вероятность $\alpha = 95$										
3	0,001									
4	0,025	0,001								
5	0,081	0,010								
6	0,146	0,034	0,004							
7	0,208	0,065	0,016							
8	0,265	0,099	0,034	0,010						
9	0,314	0,137	0,057	0,021						
10	0,356	0,172	0,083	0,037	0,014					
11	0,386	0,204	0,107	0,055	0,026					
12	0,424	0,234	0,133	0,073	0,039	0,018				
13	0,455	0,262	0,156	0,092	0,053	0,028				
14	0,484	0,293	0,179	0,112	0,068	0,039	0,021			
15	0,509	0,317	0,206	0,134	0,084	0,052	0,030			
16	0,526	0,340	0,227	0,153	0,102	0,067	0,041	0,024		
17	0,544	0,362	0,248	0,170	0,116	0,078	0,050	0,032		
18	0,562	0,382	0,267	0,187	0,132	0,091	0,062	0,041	0,026	
19	0,581	0,398	0,287	0,203	0,146	0,105	0,074	0,050	0,033	
20	0,597	0,416	0,302	0,221	0,163	0,119	0,085	0,059	0,041	0,028
25	0,652	0,493	0,381	0,298	0,236	0,186	0,146	0,114	0,089	0,068
30	0,698	0,549	0,443	0,364	0,298	0,246	0,203	0,166	0,137	0,112
35	0,732	0,596	0,495	0,417	0,351	0,298	0,254	0,214	0,171	0,154
40	0,758	0,629	0,534	0,458	0,395	0,343	0,297	0,259	0,223	0,195
45	0,778	0,658	0,567	0,492	0,433	0,381	0,337	0,299	0,263	0,233
50	0,797	0,684	0,599	0,529	0,468	0,417	0,373	0,334	0,299	0,268

Задача 298. *Для ряда наблюдений*

x_i: 0,916; 0,944; 1,292; 1,452; 1,524; 1,604; 1,632; 1,812; 2,017; 2,671

проверить наличие нескольких выбросов критерием Титьена–Мура при доверительной вероятности $\alpha = 0{,}90$.

Имеем

$$\bar{x} = \frac{1}{10} \cdot \sum_{i=1}^{10} x_i = 1{,}586; \quad \sum_{i=1}^{10} (x_i - \bar{x})^2 = 2{,}385; \quad \bar{x}_1 = \frac{1}{10-1} \cdot \sum_{i=1}^{9} x_i = 1{,}4569;$$

$$\bar{x}_2 = \frac{1}{10-2} \cdot \sum_{i=1}^{8} x_i = 1{,}397;$$

$$L_1 = \frac{\sum_{i=1}^{9} (x_i - \bar{x}_1)^2}{\sum_{i=1}^{10} (x_i - \bar{x})^2} = \frac{1{,}0787}{2{,}385} = 0{,}452; \quad L_2 = \frac{\sum_{i=1}^{8} (x_i - \bar{x}_2)^2}{\sum_{i=1}^{10} (x_i - \bar{x})^2} = \frac{0{,}737}{2{,}385} = 0{,}309.$$

Из табл. 193 для $k = 1$ $(n = 10,\ \alpha = 0{,}90)$ находим $L_1(0{,}90) = 0{,}490$, для $k = 2$ имеем $L_2(0{,}90) = 0{,}287$. Так как $L_1 = 0{,}452 < L_1(0{,}90) = 0{,}490$, наибольшее значение $x = 2{,}671$ признается выбросом. Так как $L_2 = 0{,}309 > L_2(0{,}90) = 0{,}287$, второе наибольшее значение не должно быть признано выбросом.

Для наименьших наблюдений имеем

$$\bar{x}_1^* = \frac{1}{10-1} \cdot \sum_{i=2}^{10} x_i = 1{,}661; \quad \sum_{i=2}^{10} (x_i - \bar{x}_1^*)^2 = 1{,}886; \quad \bar{x}_2^* = \frac{1}{10-2} \cdot \sum_{i=3}^{10} x_i = 1{,}750;$$

$$\sum_{i=3}^{10} (x_i - \bar{x}_2^*)^2 = 1{,}308; \quad L_1^* = \frac{1{,}886}{2{,}385} = 0{,}790; \quad L_2^* = \frac{1{,}308}{2{,}385} = 1{,}308.$$

Так как $L_1^* = 0{,}790 > L_1(0{,}90) = 0{,}490$ и $L_2^* = 0{,}548 > L_2(0{,}90) = 0{,}287$, оба наименьших наблюдения не должны признаваться выбросами.

Используем теперь E_k-критерий. Имеем $d = |x_1 - \bar{x}| = |0{,}916 - 1{,}586| = 0{,}67$.

Вычисляя далее по аналогии, получаем ранжированный ряд значений d_i. Результаты сведем в таблицу, там же укажем ранги величин x_i, соответствующие значениям d_i в ранжировке:

i	d_i	R_i	z_i	i	d_i	R_i	z_i
1	0,018	6	1,604	6	0,294	3	1,291
2	0,046	7	1,632	7	0,431	9	2,017
3	0,062	5	1,524	8	0,642	2	0,944
4	0,134	4	1,452	9	0,670	1	0,916
5	0,226	8	1,812	10	1,805	10	2,671

Пользуясь данными таблицы, вычисляем

$$\bar{z} = \frac{1}{10} \cdot \sum_{i=1}^{10} z_i = 1{,}5864; \quad \sum_{i=1}^{10} (z_i - \bar{z})^2 = 2{,}386; \quad \bar{z}_1 = \frac{1}{9} \cdot \sum_{i=1}^{9} z_i = 1{,}466;$$

$$\sum_{i=1}^{9} (z_i - \bar{z}_1)^2 = 1{,}079; \quad \bar{z}_2 = \frac{1}{8} \cdot \sum_{i=1}^{8} z_i = 1{,}535; \quad \sum_{i=1}^{8} (z_i - \bar{z}_2)^2 = 0{,}748;$$

$$E_1 = \frac{1{,}079}{2{,}386} = 0{,}452; \quad E_2 = \frac{0{,}738}{2{,}386} = 0{,}309.$$

Из табл. 194 для $\alpha = 0{,}90$ и $n = 10$ имеем $E_1(0{,}90) = 0{,}420$ и $E_2(0{,}90) = 0{,}214$.

Так как $E_1 = 0{,}452 > E_1(0{,}90) = 0{,}420$ и $E_2 = 0{,}309 > E_2(0{,}90) = 0{,}214$, оба проверяемых наблюдения (это 0,916 и 2,671) не являются выбросами.

4.3.12.1.9. Критерий Роснера для обнаружения нескольких выбросов

Рассмотренный ранее критерий Титьена–Мура (см. раздел 4.3.12.1.8) предполагает, что количество выбросов k заранее известно. Однако это не всегда так, поэтому наиболее интересны методы не только выделения самих выбросов, но и выявления их количества. Такая проблема рассмотрена Роснером [528, 529], который выдвинул и реализовал интересную идею последовательного применения критерия Груббса (см. раздел 4.3.12.1.3) для выделения одного выброса, основанного на статистике $\tau_1^* = \max\left(\dfrac{\bar{x} - x_1}{s}, \dfrac{x_n - \bar{x}}{s}\right)$.

Алгоритм критерия Роснера состоит в следующем. По начальной выборке объема n вычисляются значения $\bar{x}$ и s и статистика τ_1^*. Затем из выборки удаляется экстремальный член $x_{\min}$ $(x_{\max})$ — в зависимости от того, какое значение более удалено от среднего. Так повторяется k раз. Полученные значения статистик τ_{1i}^* $(i = 1, \ldots, k)$ каждый раз сравниваются с критическими значениями, приведенными в табл. 195 для заданных n, k и вероятности α. Превышение критерием τ_{1i}^* критического значения, полученного из табл. 195 для некоторого i, позволяет установить не только наличие выбросов, но и их количество (равное значению i, при котором появляется первая значимая величина критерия τ_{1i}^*). Вычисление последовательных статистик ведется до тех пор, пока $\tau_{1(i+1)}^* > \tau_{1i}^*$.

Задача 299. *Для ряда наблюдений* ($n = 20$)

$$x_i:\quad \begin{matrix} 0, & 15, & 16, & 22, & 22, & 23, & 26, & 27, & 27, & 28, \\ 28, & 31, & 32, & 33, & 35, & 37, & 38, & 41, & 56, & 58 \end{matrix}$$

проверить наличие выбросов и установить их количество критерием Роснера при доверительной вероятности $\alpha = 0{,}95$.

Находим для полной выборки:

$$\bar{x} = 29{,}75; \quad s = 13{,}122;$$

$$\tau_{11}^* = \max\left(\frac{x_n - \bar{x}}{s} = \frac{58 - 29{,}75}{13{,}122} = 2{,}153;\ \ \frac{\bar{x} - x_1}{s} = \frac{29{,}75 - 0}{13{,}122} = 2{,}267\right) = 2{,}267.$$

Исключаем из выборки значение $x = 0$, соответствующее полученному максимуму.

Для оставшихся членов выборки ($n = 19$) находим

$$\bar{x}_1 = 31{,}316; \quad s_1 = 11{,}402; \quad \tau_{12}^* = \max\left(\frac{58 - 31{,}316}{11{,}402} = 2{,}34;\ \ \frac{31{,}316 - 15}{11{,}402} = 1{,}431\right) = 2{,}34.$$

Так как $2{,}34 > 2{,}267$, продолжаем вычисления. Исключаем из выборки значение $x = 58$, соответствующее максимуму.

Для оставшихся $n = 18$ наблюдений имеем

$$\bar{x}_2 = 29{,}833; \quad s_2 = 9{,}666; \quad \tau_{13}^* = \max\left(\frac{56 - 29{,}833}{9{,}666} = 2{,}707;\ \ \frac{29{,}833 - 15}{9{,}666} = 1{,}534\right) = 2{,}707.$$

Так как $2{,}707 > 2{,}34$, продолжаем вычисления. Исключаем из выборки значение $x = 56$, соответствующее максимуму.

Для оставшихся $n = 17$ выборочных данных находим

$$\bar{x}_3 = 28{,}294; \quad s_3 = 7{,}346; \quad \tau_{14}^* = \max\left(\frac{41 - 28{,}294}{7{,}346} = 1{,}730;\ \ \frac{28{,}294 - 15}{7{,}346} = 1{,}810\right) = 1{,}810.$$

Так как $\tau_{14}^* = 1{,}81 < \tau_{13}^* = 2{,}707$, то дальнейшие вычисления прекращаем.

Проверим теперь значимость последовательности статистик

$$\tau_{11}^* = 2{,}267; \quad \tau_{12}^* = 2{,}34; \quad \tau_{13}^* = 2{,}707; \quad \tau_{14}^* = 1{,}81.$$

Таблица 195

Критические значения τ^*_{1i} критерия Роснера [529, 530]

n	k	i	τ^*_{1i}	n	k	i	τ^*_{1i}	n	k	i	τ^*_{1i}	n	k	i	τ^*_{1i}	n	k	i	τ^*_{1i}
Доверительная вероятность $\alpha = 90$																			
20	2	1	2,69	30	4	2	2,65	40	5	4	2,46	60	3	3	2,64	80	5	1	3,44
20	2	2	2,41	30	4	3	2,48	40	5	5	2,39	60	4	1	3,31	80	5	2	2,98
20	3	1	2,76	30	4	4	2,39	50	2	1	3,10	60	4	2	2,85	80	5	3	2,77
20	3	2	2,47	30	5	1	3,05	50	2	2	2,72	60	4	3	2,67	80	5	4	2,63
20	3	3	2,34	30	5	2	2,67	50	3	1	3,18	60	4	4	2,54	80	5	5	2,54
20	4	1	2,81	30	5	3	2,51	50	3	2	2,76	60	5	1	3,34	100	2	1	3,34
20	4	2	2,51	30	5	4	2,42	50	3	3	2,58	60	5	2	2,77	100	2	2	2,92
20	4	3	2,38	30	5	5	2,35	50	4	1	3,24	60	5	3	2,68	100	3	1	3,44
20	4	4	2,29	40	2	1	3,01	50	4	2	2,81	60	5	4	2,56	100	3	2	2,97
20	5	1	2,85	40	2	2	2,72	50	4	3	2,62	60	5	5	2,48	100	3	3	2,77
20	5	2	2,55	40	3	1	3,07	50	4	4	2,50	80	2	1	3,28	100	4	1	3,47
20	5	3	2,40	40	3	2	2,69	50	5	1	3,28	80	2	2	2,85	100	4	2	3,00
20	5	4	2,33	40	3	3	2,52	50	5	2	2,84	80	3	1	3,32	100	4	3	2,79
20	5	5	2,27	40	4	1	3,14	50	5	3	2,65	80	3	2	2,90	100	4	4	2,66
30	2	1	2,89	40	4	2	2,74	50	5	4	2,52	80	3	3	2,71	100	5	1	3,54
30	2	2	2,55	40	4	3	2,57	50	5	5	2,44	80	4	1	3,40	100	5	2	3,04
30	3	1	2,97	40	4	4	2,45	60	2	1	3,15	80	4	2	2,93	100	5	3	2,81
30	3	2	2,61	40	5	1	3,16	60	2	2	2,77	80	4	3	2,74	100	5	4	2,68
30	3	3	2,44	40	5	2	2,76	60	3	1	3,26	80	4	4	2,61	100	5	5	2,59
30	4	1	3,02	40	5	3	2,59	60	3	2	2,83								
Доверительная вероятность $\alpha = 95$																			
10	2	1	2,39	20	4	1	2,95	35	2	1	3,09	50	4	3	2,72	80	3	3	2,81
10	2	2	2,17	20	4	2	2,63	35	2	2	2,74	50	4	4	2,59	80	4	1	3,57
11	2	1	2,45	20	4	3	2,49	40	2	1	3,17	50	5	1	3,45	80	4	2	3,05
11	2	2	2,23	20	4	4	2,39	40	2	2	2,77	50	5	2	2,96	80	4	3	2,84
12	2	1	2,50	20	5	1	2,97	40	3	1	3,22	50	5	3	2,74	80	4	4	2,69
12	2	2	2,27	20	5	2	2,65	40	3	2	2,81	50	5	4	2,61	80	5	1	3,61
13	2	1	2,57	20	5	3	2,51	40	3	3	2,62	50	5	5	2,52	80	5	2	3,11
13	2	2	2,31	20	5	4	2,42	40	4	1	3,32	60	2	1	3,34	80	5	3	2,86
14	2	1	2,62	20	5	5	2,37	40	4	2	2,86	60	2	2	2,90	80	5	4	2,72
14	2	2	2,39	25	2	1	2,99	40	4	3	2,67	60	3	1	3,42	80	5	5	2,62
15	2	1	2,65	25	2	2	2,62	40	4	4	2,55	60	3	2	2,95	100	2	1	3,52
15	2	2	2,42	30	2	1	3,05	40	5	1	3,31	60	3	3	2,73	100	2	2	3,03
16	2	1	2,70	30	2	2	2,67	40	5	2	2,88	60	4	1	3,48	100	3	1	3,60
16	2	2	2,44	30	3	1	3,12	40	5	3	2,69	60	4	2	2,98	100	3	2	3,10
17	2	1	2,75	30	3	2	2,73	40	5	4	2,55	60	4	3	2,77	100	3	3	2,86
17	2	2	2,48	30	3	3	2,56	40	5	5	2,47	60	4	4	2,63	100	4	1	3,64
18	2	1	2,79	30	4	1	3,16	45	2	1	3,17	60	5	1	3,51	100	4	2	3,13
18	2	2	2,46	30	4	2	2,77	45	2	2	2,82	60	5	2	3,01	100	4	3	2,89
19	2	1	2,80	30	4	3	2,59	50	2	1	3,27	60	5	3	2,77	100	4	4	2,74
19	2	2	2,49	30	4	4	2,49	50	2	2	2,85	60	5	4	2,65	100	5	1	3,70
20	2	1	2,83	30	5	1	3,19	50	3	1	3,34	60	5	5	2,56	100	5	2	3,16
20	2	2	2,52	30	5	2	2,78	50	3	2	2,89	80	2	1	3,45	100	5	3	2,91
20	3	1	2,88	30	5	3	2,60	50	3	3	2,68	80	2	2	3,03	100	5	4	2,77
20	3	2	2,60	30	5	4	2,51	50	4	1	3,40	80	3	1	3,49	100	5	5	2,67
20	3	3	2,45	30	5	5	2,45	50	4	2	2,93	80	3	2	3,03	100			

Из табл. 195 для $\alpha = 0{,}95$, $k = 4$ и $n = 20$ находим критические значения $\tau_{11}^*(0{,}95) = 2{,}95$; $\tau_{12}(0{,}95) = 2{,}63$; $\tau_{13}(0{,}95) = 2{,}49$ и $\tau_{14}(0{,}95) = 2{,}39$.

Так как $\tau_{11}^* = 2{,}267 < \tau_{11}(0{,}95) = 2{,}95$; $\tau_{12}^* = 2{,}34 < \tau_{12}(0{,}95) = 2{,}63$; $\tau_{13}^* = 2{,}707 > \tau_{13}(0{,}95) = 2{,}49$ и $\tau_{14}^* = 1{,}81 < \tau_{14}(0{,}95) = 2{,}39$, то делаем следующий вывод: в выборке имеются $k = 3$ значимых выброса и ими являются $x = 0$, $x = 56$ и $x = 58$.

4.3.12.2. Критерии выбросов для экспоненциального распределения и распределения Вейбулла

4.3.12.2.1. Критерии выбросов для экспоненциального распределения

4.3.12.2.1.1. Критерий Смоляка–Титаренко

Для экспоненциального распределения с плотностью распределения вероятностей $f(x) = \dfrac{1}{\lambda} e^{-\frac{x}{\lambda}}$ наиболее „подозрительными" естественно считать наибольшие наблюдения, как наименее вероятные. Смоляк и Титаренко [527] рассмотрели критерий, основанный на отношении величины наибольшего члена выборки к выборочному среднему:

$$C_k = \frac{x_{n-k+1}}{\bar{x}},$$

где x_i — i-е порядковое значение ряда $x_1, x_2, \ldots, x_i, \ldots, x_n$.

Член выборки x_{n-k+1} признается выбросом при $C_k > C_k(\alpha)$. Критические значения $C_k(\alpha)$ приведены в табл. 196.

Задача 300. *Для ряда данных*

$$x_i\colon 7,\ 12,\ 15,\ 22,\ 26,\ 31,\ 34,\ 42,\ 48,\ 54,\ 61,\ 65,\ 68,\ 75,\ 80,\ 90,\ 103,\ 117,\ 132,\ 148$$

проверить наличие выбросов критерием Смоляка–Титаренко при доверительной вероятности $\alpha = 0{,}95$.

Имеем

$$\bar{x} = 61{,}5; \quad C_1 = \frac{x_n}{\bar{x}} = \frac{148}{61{,}5} = 2{,}407.$$

При $\alpha = 0{,}95$ и $n = 20$ из табл. 196 находим $C_1(0{,}95) = 5{,}408$. Так как C_1 не превышает соответствующего критического значения, то наибольшее значение не может быть признано выбросом.

4.3.12.2.1.2. Критерий Бродского–Быцаня–Власенко

Критерии, рассмотренные в работе Бродского, Быцаня и Власенко [531], являются аналогами критерия Диксона (см. раздел 4.3.12.1.6) для случая экспоненциального распределения.

Для экспоненциального распределения, записанного в форме

$$f(x) = \lambda e^{-\lambda x},$$

где λ — интенсивность отказов, статистики критериев проверки выбросов имеют вид [531]

— для проверки на выброс наибольшего значения x_n при известном λ: $z_1 = \lambda x_{n-1}$, $z_2 = \lambda(x_n - x_{n-1})$;

— для проверки на выброс наибольшего значения x_n при неизвестном λ: $z_3 = \dfrac{x_n - x_{n-1}}{x_n - x_1}$;

— для проверки на выброс наибольшего значения x_n и наименьшего значения x_1: $z_4 = \dfrac{x_n - x_{n-1}}{x_1 - x_2}$.

Таблица 196

Критические значения $C_k(\alpha)$ критерия Смоляка–Титаренко [527]

n	Доверительная вероятность α						n	Доверительная вероятность α					
	0,95			0,99				0,95			0,99		
	C_1	C_2	C_3	C_1	C_2	C_3		C_1	C_2	C_3	C_1	C_2	C_3
4	3,072			3,457			33	6,959	4,124	3,304	7,381	4,748	3,706
5	3,419			3,943			34	6,093	4,160	3,339	7,425	4,789	3,745
6	3,697			4,331			35	6,129	4,195	3,373	7,468	4,828	3,782
7	3,928			4,651			36	6,165	4,229	3,406	7,510	4,867	3,819
8	4,125	2,380		4,921	2,713		37	6,199	4,262	3,439	7,550	4,904	3,855
9	4,297	2,522		5,134	2,883		38	6,232	4,294	3,470	7,589	4,940	3,889
10	4,449	2,561		5,356	3,036		39	6,264	4,325	3,500	7,626	4,975	3,923
11	4,586	2,768		5,539	3,175		40	6,295	4,356	3,530	7,663	5,009	3,956
12	4,709	2,875	2,128	5,701	3,303	2,338	41	6,326	4,386	3,559	7,698	5,042	3,987
13	4,821	2,975	2,218	5,847	3,420	2,490	42	6,355	4,414	3,587	7,733	5,074	4,018
14	4,924	3,067	2,303	5,981	3,528	2,585	43	6,384	4,442	3,615	7,766	5,106	4,048
15	5,019	3,152	2,382	6,013	3,629	2,765	44	6,412	4,470	3,641	7,799	5,137	4,078
16	5,107	3,232	2,456	6,126	3,723	2,758	45	6,439	4,497	3,668	7,831	5,167	4,107
17	5,190	3,308	2,526	6,321	3,811	2,837	46	6,467	4,523	3,694	7,861	5,196	4,135
18	5,267	3,378	2,593	6,418	3,893	2,911	47	6,492	4,548	3,719	7,892	5,224	4,162
19	5,340	3,445	2,656	6,509	3,971	2,987	48	6,518	4,574	3,743	7,921	5,252	4,189
20	5,408	3,509	2,716	6,594	4,045	3,050	49	6,543	4,598	3,767	7,950	5,279	4,225
21	5,473	3,570	2,773	6,674	4,114	3,114	50	6,567	4,622	3,791	7,978	5,306	4,241
22	5,553	3,627	2,827	6,750	4,181	3,175	55	6,682	4,734	3,901	8,108	5,430	4,362
23	5,592	3,682	2,879	6,822	4,244	3,233	60	6,785	4,837	4,002	8,226	5,542	4,472
24	5,648	3,734	2,929	6,890	4,304	3,289	65	6,880	4,930	4,094	8,332	5,645	4,572
25	5,701	3,784	2,977	6,955	4,362	3,343	70	6,967	5,016	4,179	8,430	5,739	4,664
26	5,752	3,833	3,028	7,016	4,417	3,395	75	7,047	5,096	4,258	8,519	5,826	4,749
27	5,800	3,879	3,068	7,075	4,470	3,444	80	7,122	5,170	4,332	8,602	5,906	4,829
28	5,847	3,923	3,111	7,132	4,521	3,492	85	7,192	5,239	4,401	8,680	5,982	4,903
29	5,892	3,966	3,152	7,186	4,570	3,538	90	7,257	5,305	4,466	8,752	6,052	4,973
30	5,935	4,008	3,192	7,237	4,617	3,582	95	7,319	5,367	4,527	8,820	6,119	5,038
31	5,977	4,048	3,230	7,287	4,662	3,625	100	7,378	5,425	4,584	8,884	6,181	5,098
32	6,020	4,086	3,268	7,335	4,706	3,667							

Для статистик перечисленных критериев получены полезные результаты [531]:

$$\mathbf{P}(z_1 \geqslant x) = 1 - \left\{1 - e^{-(1+x)}\right\}^n; \quad \mathbf{P}(z_3 \geqslant x) = (n-1)!(1-x)^{n-2} \prod_{j=1}^{n-2} \frac{1}{1+j(1-x)};$$

$$\mathbf{P}(z_2 \geqslant x) = e^{-x}; \quad \mathbf{P}(z_4 \geqslant x) = (n-2)!(1-x)^{n-3} \prod_{j=1}^{n-3} \frac{1}{1+j(1-x)}.$$

Если принятый уровень значимости α меньше, чем любая из указанных вероятностей (в зависимости от используемого критерия), гипотеза о наличии выбросов отклоняется.

Задача 301. *Для данных задачи* 300 *проверить наличие выбросов критериями* z_3 *и* z_4 *настоящего раздела на уровне значимости* $\alpha = 0{,}05$.

Имеем $z_3 = \dfrac{148-132}{148-7} = 0{,}113$; $z_4 = \dfrac{148-132}{148-12} = 0{,}1176$. Вычисляем $\mathbf{P}(z_3 \geqslant 0{,}113) = (20-1)! \cdot (1-0{,}113)^{18} \cdot \prod_{j=1}^{18} \dfrac{1}{1+j\cdot(1-0{,}113)} = 0{,}7262$.

Так как $\mathbf{P}(z_3 \geqslant 0{,}113) = 0{,}726 > \alpha = 0{,}05$, то наибольшее значение не является выбросом.

Аналогично для $z_4 = 0{,}1176$ имеем $\mathbf{P}(z_4 \geqslant 0{,}1176) = 0{,}7207$, что также больше 0,05 и, следовательно, также отклоняет гипотезу наличия выброса.

4.3.12.2.1.3. Критерий Кимбера для нескольких выбросов

Кимбер [532] предложил по аналогии с критерием Роснера (см. раздел 4.3.12.1.9) последовательную процедуру для выявления нескольких выбросов в выборке из экспоненциального распределения. Статистика критерия Кимбера S_j для выделения j наибольших выбросов имеет вид [532]

$$S_j = \frac{x_{n-j+1}}{\sum\limits_{i=1}^{n-j+1} x_i}, \quad j = 1, 2, \ldots, n-1.$$

Если $S_j(\alpha)$ — критическое значение статистики, то правило принятия решения по гипотезе включает в себя следующую последовательность операций. Если для $i = 1, 2, \ldots, k$ имеет место $S_i < S_i(\alpha)$, то с достоверностью α признается, что k наибольших наблюдений в выборке не являются выбросами. Если $S_i > S_i(\alpha)$ для $i = k, k-1, \ldots, j+1$ и $S_j > S_j(\alpha)$, то j наибольших наблюдений являются выбросами. Если $S_k > S_k(\alpha)$, то k наибольших наблюдений являются выбросами.

Для выделения нижних выбросов используется статистика

$$S_j^* = \frac{x_{j+1}}{\sum\limits_{i=1}^{j+1} x_i}, \quad j = 1, 2, \ldots, n-1.$$

Очевидно, что $S_j^* = S_{n-j}$. Процедура выделения k нижних выбросов аналогична процедуре выделения k верхних выбросов, но с заменой $S_k, S_{k-1}, \ldots, S_1$ на $S_k^*, S_{k-1}^*, \ldots, S_1^*$. Критические значения $S_j(\alpha)$ для выделения нижних и верхних выбросов приведены в табл. 197 и 198.

Таблица 197

Критические значения $S_j(\alpha)$ статистики Кимбера для выделения j верхних выбросов [532]

n	Доверительная вероятность α							
	0,95				0,99			
	j				j			
	1	2	3	4	1	2	3	4
$k=2$								
10	0,4830	0,4340			0,5700	0,5140		
11	0,4530	0,4010			0,5360	0,4750		
12	0,4270	0,3720			0,5060	0,4410		
13	0,4030	0,3480			0,4800	0,4120		
14	0,3830	0,3270			0,4560	0,3870		
15	0,3640	0,3080			0,4350	0,3640		
16	0,3470	0,2920			0,4150	0,3450		
17	0,3320	0,2770			0,3980	0,3270		
18	0,3180	0,2640			0,3810	0,3120		
19	0,3060	0,2520			0,3660	0,2970		
20	0,2940	0,2410			0,3530	0,2840		
22	0,2730	0,2220			0,3280	0,2620		
24	0,2560	0,2060			0,3070	0,2420		
26	0,2400	0,1930			0,2890	0,2260		
28	0,2270	0,1810			0,2730	0,2120		
30	0,2150	0,1710			0,2580	0,2000		
35	0,1900	0,1500			0,2290	0,1750		
40	0,1710	0,1340			0,2050	0,1560		
45	0,1550	0,1210			0,1860	0,1400		
50	0,1420	0,1110			0,1710	0,1280		
60	0,1220	0,0470			0,1470	0,1090		
70	0,1070	0,0630			0,1290	0,0955		
80	0,0960	0,0740			0,1150	0,0885		
90	0,0869	0,0670			0,1040	0,0767		
100	0,0894	0,0612			0,0948	0,0699		
120	0,0679	0,0523			0,0809	0,0597		
140	0,0595	0,0459			0,0707	0,0521		
$k=3$								
15	0,3800	0,3210	0,3060		0,4520	0,3790	0,3590	
16	0,3630	0,3040	0,2870		0,4320	0,3590	0,0337	
17	0,3470	0,2880	0,2710		0,4120	0,3390	0,3160	
18	0,3330	0,2740	0,2570		0,3950	0,3220	0,2990	
19	0,3250	0,2660	0,2480		0,3800	0,3070	0,2840	
20	0,3130	0,2550	0,2360		0,3660	0,2940	0,2700	
22	0,2910	0,2350	0,2160		0,3430	0,2720	0,2480	
24	0,2690	0,2140	0,1960		0,3210	0,2520	0,2280	
26	0,2510	0,2000	0,1830		0,3020	0,2350	0,2120	
28	0,2360	0,1880	0,1700		0,2830	0,2190	0,1960	
30	0,2290	0,1800	0,1630		0,2680	0,2060	0,1840	
35	0,1980	0,1550	0,1390		0,2390	0,1820	0,1600	
40	0,1790	0,1400	0,1250		0,2130	0,1600	0,1410	
45	0,1660	0,1280	0,1130		0,1930	0,1450	0,1260	
50	0,1470	0,1140	0,1010		0,1790	0,1330	0,1160	
60	0,1270	0,0973	0,0860		0,1540	0,1130	0,0983	

Окончание таблицы 197

n	Доверительная вероятность α							
	0,95				0,99			
	j				j			
	1	2	3	4	1	2	3	4
$k = 3$								
70	0,1110	0,0852	0,0751		0,1330	0,0981	0,0849	
80	0,1030	0,0781	0,0685		0,1190	0,0873	0,0754	
90	0,0932	0,0706	0,0706		0,1090	0,0796	0,9685	
100	0,0853	0,0636	0,0559		0,0994	0,0726	0,0624	
120	0,0626	0,0476	0,0417		0,0727	0,0532	0,0458	
$k = 4$								
20	0,3210	0,2610	0,2410	0,2350	0,3770	0,3020	0,2770	0,2690
22	0,2990	0,2400	0,2210	0,2130	0,3510	0,2780	0,2530	0,2430
24	0,2800	0,2230	0,2030	0,1950	0,3290	0,2580	0,2330	0,2220
26	0,2630	0,2080	0,1890	0,1800	0,3090	0,2400	0,2160	0,2050
28	0,2480	0,1950	0,1760	0,1680	0,2920	0,2250	0,2010	0,1900
30	0,2350	0,1840	0,1660	0,1570	0,2770	0,2120	0,1890	0,1770
35	0,2080	0,1610	0,1440	0,1360	0,2450	0,1850	0,1630	0,1520
40	0,1870	0,1440	0,1280	0,1200	0,2200	0,1650	0,1440	0,1340
45	0,1690	0,1300	0,1150	0,1070	0,2000	0,1490	0,1300	0,1200
50	0,1550	0,1190	0,1050	0,0974	0,1830	0,1360	0,1180	0,1080
60	0,1340	0,1010	0,0892	0,0826	0,1570	0,1115	0,0998	0,0915
70	0,1170	0,0888	0,0779	0,0719	0,1380	0,1010	0,0869	0,0795
80	0,1050	0,0792	0,0693	0,0639	0,1230	0,0897	0,0772	0,0704
90	0,0948	0,0715	0,0625	0,0576	0,1110	0,0809	0,0695	0,0633
100	0,0866	0,0653	0,0571	0,0525	0,1010	0,0737	0,0633	0,0576
120	0,0740	0,0588	0,0487	0,0447	0,0866	0,0628	0,0538	0,0489
140	0,0647	0,0488	0,0426	0,0391	0,0756	0,0546	0,0470	0,0270

Задача 302. *Для ряда данных задачи 300 проверить наличие выбросов критерием Кимбера при доверительной вероятности* $\alpha = 0{,}95$.

Вычисляем

$$S_1 = \frac{x_{20}}{\sum\limits_{i=1}^{20} x_i} = \frac{148}{1230} = 0{,}120; \quad S_2 = \frac{x_{19}}{\sum\limits_{i=1}^{19} x_i} = \frac{132}{1082} = 0{,}122;$$

$$S_3 = \frac{x_{18}}{\sum\limits_{i=1}^{18} x_i} = \frac{117}{950} = 0{,}123; \quad S_4 = \frac{x_{17}}{\sum\limits_{i=1}^{17} x_i} = \frac{103}{833} = 0{,}124.$$

Используя табл. 197 для $n = 20$, $\alpha = 0{,}95$, $k = 4$, убеждаемся, что $S_4 = 0{,}124 < S_4(0{,}95) = 0{,}235$, и по аналогии $S_3 < S_3(0{,}95)$, $S_2 < S_2(0{,}95)$, $S_1 < S_1(0{,}95)$. Следовательно, верхних выбросов в выборке нет.

Проверяем нижние выбросы. Имеем

$$S_1^* = \frac{x_2}{x_1 + x_2} = \frac{12}{19} = 0{,}632; \quad S_2^* = \frac{x_3}{x_1 + x_2 + x_3} = \frac{15}{34} = 0{,}441;$$

$$S_3^* = \frac{x_4}{\sum\limits_{i=1}^{4} x_i} = \frac{22}{56} = 0{,}393; \quad S_4^* = \frac{x_5}{\sum\limits_{i=1}^{5} x_i} = \frac{26}{82} = 0{,}317.$$

Обратившись к табл. 198 ($k=4$), убеждаемся, что

$$S_1^* > S_1^*(0{,}95) = 0{,}596; \quad S_2^* < S_2^*(0{,}95) = 0{,}717;$$
$$S_3^* < S_3^*(0{,}95) = 0{,}868; \quad S_4^* < S_4^*(0{,}95) = 0{,}988.$$

Следовательно, только первое нижнее значение ($x_1 = 7$) является выбросом.

Таблица 198

Критические значения $S_j^*(\alpha)$ критерия Кимбера для выделения j нижних выбросов [532]

n	Доверительная вероятность α							
	0,95				0,99			
	j				j			
	1	2	3	4	1	2	3	4
$k=2$								
10	0,977	0,837			0,996	0,921		
15	0,977	0,829			0,995	0,917		
20	0,976	0,825			0,995	0,915		
50	0,975	0,819			0,995	0,911		
100	0,975	0,817			0,995	0,910		
200	0,975	0,816			0,995	0,910		
$k=3$								
15	0,984	0,855	0,705		0,997	0,931	0,808	
20	0,984	0,852	0,698		0,997	0,929	0,802	
50	0,983	0,846	0,687		0,997	0,926	0,792	
100	0,983	0,845	0,683		0,997	0,925	0,789	
200	0,983	0,843	0,681		0,997	0,925	0,788	
$k=4$								
20	0,988	0,868	0,717	0,596	0,998	0,938	0,817	0,693
50	0,987	0,863	0,706	0,580	0,998	0,935	0,807	0,677
10	0,987	0,861	0,702	0,575	0,997	0,934	0,804	0,672
200	0,987	0,860	0,700	0,573	0,997	0,934	0,803	0,671

4.3.12.2.2. Критерии выбросов для распределения Вейбулла

В [522, 534–537] представлен достаточно полный набор различных методов выявления выбросов в выборках, большинство из которых рассмотрено выше. В [533] рассмотрены методы трансформации статистик известных критериев для обнаружения выбросов в выборках, извлекаемых из генеральных совокупностей, имеющих распределение Вейбулла.

Для выделения верхних выбросов статистиками критериев являются:

— критерии типа Груббса [522]

$$G = \frac{S^2_{n-k+1,\ldots,n}}{S^2}, \quad \text{где} \quad S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2; \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{nk} x_i;$$

$$S^2_{n-k+1,\ldots,n} = \sum_{i=1}^{n-k}(x_i - \bar{x}_{n-k+1,\ldots,n})^2; \quad \bar{x}_{n-k+1,\ldots,n} = \frac{1}{n-k}\sum_{i=1}^{n-k} x_i;$$

— критерии типа Диксона [524, 636]

$$R_1 = \frac{x_n - x_{n-k}}{x_n - x_1}; \quad R_2 = \frac{x_n - x_{n-k}}{x_n - x_2}; \quad R_3 = \frac{x_n - x_{n-k}}{x_n - x_3}.$$

Для выделения нижних выбросов статистиками критериев являются

— критерии типа Груббса

$$G^* = \frac{S^2_{1,\ldots,k}}{S^2}, \quad \text{где} \quad S^2_{1,\ldots,k} = \sum_{i=k+1}^{n} \left(x_i - \bar{x}_{1,\ldots,k}\right)^2; \quad \bar{x}_{1,\ldots,k} = \frac{1}{n-k} \sum_{i=k+1}^{n} x_i;$$

— критерии типа Диксона

$$R_1^* = \frac{x_{k+1} - x_1}{x_n - x_1}; \quad R_2^* = \frac{x_{k+1} - x_1}{x_{n-1} - x_1}; \quad R^* = \frac{x_{k+1} - x_1}{x_{n-2} - x_1}.$$

Критические значения $G(\alpha)$ и $R_i(\alpha)$ приведены в табл. 199, $G_k^*(\alpha)$ и $R_i^*(\alpha)$ — в табл. 200.

При $G < G(\alpha)$ и $R_k > R_k(\alpha)$ в выборке признается наличие k верхних выбросов. При $G_k^* < G_k^*(\alpha)$ и $R_k^* > R_k^*(\alpha)$ — признается наличие k нижних выбросов.

Задача 303. *Для выборки наблюдений* ($n = 20$)

x_i:									
2	3	11	14	19	25	26	31	39	41
46	54	59	65	71	81	95	120	138	154

проверить наличие в выборке выбросов критериями Груббса и Диксона для распределения Вейбулла при доверительной вероятности $\alpha = 0{,}95$.

В нашем случае $n = 20$, $x_n = 154$, $x_1 = 2$.

Результаты вычислений сводим в таблицу, используя критические значения из табл. 199 и 200 при $n = 20$, $k = 1, 2, 3$ и $\alpha = 0{,}95$:

Статистика	k			Статистика	k		
	1	2	3		1	2	3
$\bar{x}_{n-k+1}$	49,474	44,55	40,12	$\bar{x}_k$	57,47	60,5	63,42
S^2_{n-k+1}	26459	18186	12160	S^2_k	33915	30782	28188
x_{n-k}	138	120	95	x_{k+1}	3	11	14
R_1	0,105	0,224	0,388	R_1^*	0,0066	0,059	0,079
$R_1(0{,}95)$	0,18	0,24	0,29	$R_1^*(0{,}95)$	0,46	0,56	0,61
R_2	0,106	0,225	0,391	R_2^*	0,0073	0,066	0,088
$R_2(0{,}95)$	0,21	0,29	0,34	$R_2^*(0{,}95)$	0,49	0,59	0,64
R_3	0,112	0,238	0,412	R_3^*	0,0085	0,076	0,1016
$R_3(0{,}95)$	0,24	0,29	0,34	$R_3^*(0{,}95)$	0,51	0,61	0,67
G_k	0,718	0,494	0,33	G_k^*	0,921	0,836	0,765
$G_k(0{,}95)$	0,78	0,65	0,54	$G_k(0{,}95)$	0,41	0,27	0,19

Из таблицы следует, что три наибольших наблюдения $x_i = 120$, 138 и 154 признаются выбросами, так как для них $G_3 = 0{,}330 < G_3(0{,}95) = 0{,}54$ и $R_1 = 0{,}388 > R_1(0{,}95) = 0{,}29$, $R_2 = 0{,}391 > R_2(0{,}95) = 0{,}34$, $R_3 = 0{,}412 > R_3(0{,}95) = 0{,}34$. Так как $G_k^* > G_k^*(\alpha)$, то нижние выбросы в выборке отсутствуют.

4.3.12.3. Критерий выбросов для любого непрерывного распределения (критерий Дарлинга)

Для любого непрерывного распределения, определенного с точностью до параметров, Дарлинг [538] предложил критерий, основанный на статистике $l_n = \sum_{i=1}^{n} \frac{F(x_i)}{F(x_n)}$ $\left(\sum_{i=1}^{n} \frac{-F(x_i)}{1 - F(x_n)}\right)$ для проверки на выброс наибольшего (наименьшего) наблюдения.

Таблица 199

Критические значения $G(\alpha)$ и $R_i(\alpha)$ критериев типа Груббса и Диксона для выявления k верхних выбросов в выборке, имеющей распределение Вейбулла [533]

n	Доверительная вероятность α							
	0,90				0,95			
	G	R_1	R_2	R_3	G	R_1	R_2	R_3
$k = 1$								
5	0,30	0,46	0,67	0,90	0,20	0,55	0,76	0,95
6	0,41	0,38	0,53	0,70	0,31	0,46	0,62	0,79
7	0,49	0,32	0,44	0,57	0,41	0,38	0,52	0,66
8	0,55	0,28	0,38	0,48	0,48	0,34	0,45	0,56
9	0,60	0,26	0,34	0,42	0,53	0,32	0,40	0,49
10	0,64	0,24	0,31	0,37	0,58	0,29	0,37	0,44
11	0,67	0,22	0,28	0,34	0,61	0,27	0,34	0,40
12	0,69	0,20	0,26	0,31	0,65	0,25	0,32	0,37
13	0,71	0,19	0,25	0,29	0,67	0,24	0,30	0,35
14	0,74	0,18	0,23	0,27	0,69	0,22	0,28	0,32
15	0,75	0,17	0,22	0,25	0,71	0,21	0,26	0,30
16	0,77	0,17	0,21	0,24	0,73	0,21	0,25	0,29
17	0,78	0,16	0,20	0,23	0,75	0,20	0,24	0,28
18	0,79	0,16	0,19	0,22	0,76	0,19	0,23	0,26
19	0,80	0,15	0,18	0,21	0,77	0,18	0,22	0,25
20	0,81	0,14	0,18	0,20	0,78	0,18	0,21	0,24
$k = 2$								
8	0,32	0,43	0,56	0,68	0,25	0,49	0,62	0,75
9	0,38	0,39	0,49	0,59	0,31	0,44	0,56	0,66
10	0,42	0,35	0,44	0,52	0,37	0,40	0,50	0,59
11	0,47	0,32	0,40	0,48	0,41	0,37	0,46	0,54
12	0,50	0,29	0,37	0,44	0,45	0,34	0,43	0,50
13	0,54	0,28	0,35	0,40	0,48	0,32	0,40	0,46
14	0,56	0,26	0,32	0,38	0,51	0,31	0,37	0,43
15	0,59	0,25	0,31	0,36	0,55	0,29	0,35	0,40
16	0,61	0,24	0,29	0,34	0,57	0,28	0,34	0,38
17	0,63	0,23	0,28	0,32	0,59	0,27	0,32	0,36
18	0,65	0,22	0,27	0,31	0,62	0,26	0,31	0,35
19	0,67	0,21	0,26	0,29	0,63	0,25	0,30	0,34
20	0,69	0,20	0,25	0,28	0,65	0,24	0,29	0,32
$k = 3$								
10	0,29	0,44	0,56	0,66	0,24	0,50	0,62	0,72
12	0,38	0,37	0,46	0,54	0,32	0,42	0,52	0,60
14	0,45	0,33	0,40	0,46	0,39	0,37	0,45	0,51
16	0,50	0,29	0,36	0,41	0,45	0,34	0,40	0,46
18	0,55	0,27	0,33	0,37	0,50	0,31	0,37	0,41
20	0,58	0,25	0,30	0,34	0,54	0,29	0,34	0,38

Здесь $F(x_i) = \mathbf{P}(x < x_i)$ — функция распределения вероятностей в точке $x = x_i$.

При $n > 3$ распределение l_n близко к нормальному со средним $\mathbf{M}(l_n)$ и дисперсией $\mathbf{D}(l_n)$, где

$$\mathbf{M}(l_n) = \frac{n+1}{2}; \quad \mathbf{D}(l_n) = \frac{n-1}{12},$$

Таблица 200

Критические значения $G^*(\alpha)$ и $R_i^*(\alpha)$ критериев типа Груббса и Диксона для выявления k нижних выбросов в выборке, имеющей распределение Вейбулла [533]

n	Доверительная вероятность α							
	0,90				0,95			
	G^*	R_1^*	R_2^*	R_3^*	G^*	R_1^*	R_2^*	R_3^*
$k=1$								
5	0,11	0,67	0,79	0,94	0,07	0,74	0,86	0,97
6	0,16	0,60	0,70	0,80	0,10	0,68	0,77	0,87
7	0,20	0,56	0,64	0,72	0,14	0,64	0,71	0,79
8	0,25	0,53	0,60	0,66	0,18	0,60	0,67	0,74
9	0,28	0,51	0,57	0,61	0,20	0,58	0,64	0,69
10	0,31	0,49	0,54	0,58	0,24	0,56	0,61	0,65
11	0,34	0,47	0,52	0,55	0,26	0,55	0,59	0,62
12	0,36	0,46	0,49	0,53	0,28	0,53	0,56	0,61
13	0,38	0,45	0,49	0,51	0,30	0,52	0,56	0,58
14	0,40	0,44	0,47	0,50	0,32	0,51	0,54	0,57
15	0,42	0,43	0,46	0,49	0,34	0,50	0,53	0,56
16	0,43	0,43	0,45	0,47	0,35	0,50	0,52	0,55
17	0,45	0,42	0,44	0,46	0,37	0,49	0,51	0,53
18	0,46	0,41	0,43	0,46	0,38	0,48	0,50	0,53
19	0,48	0,40	0,43	0,44	0,40	0,47	0,50	0,51
20	0,49	0,39	0,42	0,44	0,41	0,46	0,49	0,51
$k=2$								
8	0,10	0,67	0,75	0,82	0,07	0,73	0,80	0,86
9	0,13	0,64	0,71	0,76	0,09	0,70	0,76	0,81
10	0,15	0,62	0,68	0,72	0,11	0,68	0,73	0,77
11	0,18	0,60	0,65	0,69	0,13	0,65	0,71	0,75
12	0,20	0,58	0,63	0,66	0,15	0,64	0,68	0,72
13	0,22	0,57	0,61	0,64	0,17	0,62	0,67	0,70
14	0,23	0,55	0,60	0,62	0,18	0,61	0,65	0,68
15	0,25	0,54	0,58	0,61	0,20	0,60	0,64	0,66
16	0,27	0,53	0,56	0,59	0,22	0,59	0,62	0,65
17	0,28	0,52	0,56	0,58	0,23	0,58	0,61	0,64
18	0,30	0,52	0,55	0,57	0,25	0,57	0,60	0,63
19	0,31	0,51	0,54	0,56	0,26	0,57	0,59	0,62
20	0,32	0,50	0,53	0,55	0,27	0,56	0,59	0,61
$k=3$								
10	0,08	0,70	0,76	0,82	0,06	0,75	0,80	0,86
12	0,12	0,66	0,71	0,75	0,09	0,71	0,75	0,79
14	0,15	0,63	0,67	0,70	0,12	0,68	0,72	0,74
16	0,18	0,61	0,64	0,67	0,14	0,65	0,68	0,71
18	0,21	0,58	0,61	0,64	0,17	0,63	0,66	0,69
20	0,23	0,57	0,60	0,62	0,19	0,61	0,64	0,67

т. е. если

$$l_n < \sqrt{\frac{n-1}{12}}u_{1-\alpha} + \frac{n+1}{2} \qquad \left(l_n > \sqrt{\frac{n-1}{12}}u_{\alpha} + \frac{n+1}{2}\right),$$

то с вероятностью α нижнее (верхнее) выборочное значение признается выбросом (u_α — α-квантиль стандартного нормального распределения).

В [539, 540] предложена модификация критерия Дарлинга, имеющая бóльшую мощность, чем исходный критерий. Предложены статистики:
— для проверки на выброс наибольшего значения x_n

$$L = -2\sum_{i=1}^{n-1} \ln\left[1 - \frac{F(x_i)}{F(x_n)}\right];$$

— для проверки на выброс наименьшего значения x_1

$$L^* = -2\sum_{i=1}^{n-1} \ln\left[1 - \frac{1-F(x_{i+1})}{1-F(x_1)}\right].$$

При $L\ (L^*) \geqslant \chi^2_\alpha[2(n-1)]$ в выборке отсутствуют выбросы, при $L\ (L^*) < \chi^2_\alpha[2(n-1)]$ x_n (x_1) является выбросом. Здесь $\chi^2_\alpha[2(n-1)]$ — α-квантиль χ^2-распределения с $f = 2\ (n-1)$ степенями свободы.

При нормальном исходном распределении мощность этого критерия по сравнению с критерием Груббса падает не более, чем на 20%.

Задача 304. *Для ряда значений выборки, взятой из экспоненциального распределения с плотностью* $f(x) = \frac{1}{\lambda}\exp\left(-\frac{x}{\lambda}\right)$ *и параметром* $\lambda = 21{,}5$:

$$x_i\text{: } 1,\ 4,\ 9,\ 16,\ 23,\ 24,\ 28,\ 31,\ 56,\ 63$$

проверить наличие выбросов критерием Дарлинга при доверительной вероятности $\alpha = 0{,}95$.

Функция нашего распределения $F(x) = 1 - \exp\left(-\frac{x}{21{,}5}\right)$.

Находим

$$F(x_1) = 1 - \exp\left(-\frac{1}{21{,}5}\right) = 0.0454; \quad F(x_2) = 1 - \exp\left(-\frac{4}{21{,}5}\right) = 0{,}1627;$$

$$F(x_3) = 0.342; \quad F(x_4) = 0.5249; \quad F(x_5) = 0{,}6569; \quad F(x_6) = 0{,}67256;$$

$$F(x_7) = 0.7281; \quad F(x_8) = 0.7635; \quad F(x_9) = 0{,}9251; \quad F(x_{10}) = 0.9466.$$

Далее вычисляем

$$L = -2\cdot\sum_{i=1}^{9} \ln\left[1 - \frac{F(x_i)}{F(x_n)}\right] = -2\cdot\left[\ln\left(1 - \frac{0{,}0454}{0{,}9466}\right) + \ldots + \ln\left(1 - \frac{0{,}9261}{0{,}9466}\right)\right] = 21{,}736.$$

Из табл. 55 находим $\chi^2_{0,95}(2\cdot 9) = 28.869$. Так как $L = 21{,}736 < 28.869$, то с вероятностью 0.95 наибольшее значение в выборке $x_n = 63$ следует считать выбросом.

Вычислим теперь

$$L^* = -2\cdot\sum_{i=1}^{9} \ln\left[1 - \frac{-F(x_{i+1})}{F(x_1)}\right] = -2\cdot(-2{,}03859 - 1{,}1689 - \ldots - 0{,}05756) = 11.4622.$$

Сравнением $L^* = 11{,}46 < \chi^2_{0,95}(18) = 28.869$ убеждаемся, что наименьшее значение $x_1 = 1$ также является выбросом.

4.4. Толерантные пределы

Ценную в практическом отношении информацию о распределении наблюдаемой случайной величины дают границы интервала, в котором с заданной вероятностью α будет находиться β-я часть совокупности значений случайной величины. Такие интервалы (или пределы, в случае односторонних интервалов) называются толерантными.

Например, широко используемый показатель при нормировании характеристик надежности технических изделий — γ-ресурс, определяемый значением наработки изделия, при котором доля дефектных изделий в партии не будет превышать $(1-\gamma)$, в математической статистике интерпретируется как односторонний толерантный предел при $\gamma = \beta$.

4.4.1. Толерантные пределы в случае нормального распределения

4.4.1.1. Толерантные пределы при известных параметрах распределения (μ и σ^2)

Двусторонний симметричный относительно μ интервал имеет вид

$$\left[\mu - u_{\frac{1+\beta}{2}}\sigma;\ \mu + u_{\frac{1+\beta}{2}}\sigma\right].$$

Односторонние толерантные интервалы равны

$$(-\infty; \mu + u_\beta\sigma] \quad \text{и} \quad [\mu - u_\beta\sigma; \infty)\,,$$

где u_ε — ε-квантиль стандартного нормального распределения.

Задача 305. *Найти двусторонний толерантный интервал, в котором будет находиться 75% всей совокупности значений нормально распределенной случайной величины с параметрами распределения $\mu = 101$ и $\sigma = 24{,}8$.*

В нашем случае $\beta = 0{,}75$ и

$$u_{\frac{1+0{,}75}{2}} = u_{0{,}875} \approx 4{,}91\cdot\left(0{,}875^{0{,}14} - 0{,}125^{0{,}14}\right) = 1{,}149.$$

Тогда искомый интервал ограничен величинами

$$101 - 1{,}149\cdot 24{,}8 = 72{,}5 \quad \text{и} \quad 101 + 1{,}149\cdot 24{,}8 = 129{,}5.$$

Таким образом, в интервале от 72,5 до 129,5 будет находиться 75% всех значений случайной величины.

4.4.1.2. Толерантные пределы при неизвестных параметрах распределения

4.4.1.2.1. Среднее μ неизвестно, дисперсия σ^2 известна

Для односторонних толерантных интервалов $(-\infty; \tau_{\text{в}}]$ и $[\tau_{\text{н}}; \infty)$, где $\tau_{\text{в}} = \bar{x} + a(n,\alpha,\beta)\sigma$ и $\tau_{\text{н}} = \bar{x} - a(n,\alpha,\beta)\sigma$, справедливо утверждение о том, что β-я часть совокупности значений случайной величины x с вероятностью α находится в каждом из них.

Множитель $a(n,\alpha,\beta)$ является решением уравнения

$$\Phi\left(-\frac{u_\alpha}{\sqrt{n}} + a(n,\alpha,\beta)\right) = \beta,$$

откуда следует, что $a(n,\alpha,\beta) = u_\beta + \dfrac{u_\alpha}{\sqrt{n}}$, где u_γ — γ-квантиль стандартного нормального распределения, Φ — функция Лапласа.

Таблица 201

Множители $a(n, \alpha, \beta)$ для построения односторонних толерантных пределов в случае нормального распределения (μ неизвестно, σ^2 известно) [56]

n	α					
	0,90			0,95		
	β			β		
	0,90	0,95	0,99	0,90	0,95	0,99
1	2,563	2,926	3,608	2,926	3,290	3,971
2	2,188	2,551	3,233	2,445	2,808	3,489
3	2,021	2,385	3,066	2,231	2,595	3,276
4	1,922	2,286	2,967	2,104	2,467	3,149
5	1,855	2,218	2,899	2,017	2,380	3,062
6	1,805	2,168	2,850	1,953	2,316	2,998
7	1,766	2,129	2,811	1,903	2,267	2,948
8	1,735	2,098	2,779	1,863	2,226	2,908
9	1,709	2,072	2,754	1,830	2,193	2,875
10	1,687	2,050	2,732	1,802	2,165	2,846
11	1,668	2,031	2,713	1,777	2,141	2,822
12	1,652	2,015	2,695	1,756	2,120	2,801
13	1,637	2,000	2,682	1,738	2,101	2,783
14	1,624	1,987	2,669	1,721	2,084	2,766
15	1,612	1,976	2,657	1,706	2,070	2,751
16	1,602	1,965	2,647	1,693	2,056	2,738
17	1,592	1,956	2,637	1,680	2,044	2,725
18	1,584	1,947	2,628	1,669	2,033	2,714
19	1,576	1,939	2,620	1,659	2,022	2,704
20	1,568	1,931	2,613	1,649	2,013	2,694
25	1,538	1,901	2,583	1,611	1,974	2,655
30	1,516	1,879	2,560	1,582	1,945	2,627
35	1,498	1,861	2,543	1,560	1,923	2,604
40	1,484	1,847	2,529	1,542	1,905	2,586
45	1,473	1,836	2,517	1,527	1,890	2,572
50	1,463	1,826	2,508	1,514	1,877	2,559
60	1,447	1,810	2,492	1,494	1,857	2,539
70	1,435	1,798	2,480	1,478	1,841	2,523
80	1,425	1,788	2,470	1,465	1,829	2,510
90	1,417	1,780	2,461	1,455	1,818	2,500
100	1,410	1,773	2,455	1,446	1,809	2,491
200	1,372	1,735	2,417	1,398	1,761	2,443
300	1,356	1,719	2,400	1,377	1,740	2,421
400	1,346	1,709	2,390	1,364	1,727	2,409
500	1,339	1,702	2,384	1,355	1,718	2,400
1000	1,322	1,685	2,367	1,334	1,697	2,378
∞	1,282	1,645	2,326	1,282	1,645	2,326

Значения множителей $a(n, \alpha, \beta)$ приведены в табл. 201.

Для двусторонних толерантных интервалов $(\tau_в, \tau_н)$ справедливы формулы

$$\tau_в = \bar{x} + a^*(n, \alpha, \beta)\,\sigma \quad \text{и} \quad \tau_н = \bar{x} - a^*(n, \alpha, \beta)\,\sigma.$$

Для них справедливо утверждение, что с вероятностью α β-я часть распределения случайной величины x заключена в промежутке $[\tau_н; \tau_в]$. Значения множителей

Таблица 202

Множители $a^*(n, \alpha, \beta)$ для построения двусторонних толерантных интервалов в случае нормального распределения (μ неизвестно, σ^2 известно) [56]

n	α					
	0,90			0,95		
	β			β		
	0,90	0,95	0,99	0,90	0,95	0,99
1	2,927	3,290	3,972	3,242	3,605	4,287
2	2,446	2,809	3,490	2,668	3,031	3,713
3	2,236	2,597	3,277	2,415	2,778	3,459
4	2,114	2,473	3,152	2,265	2,627	3,308
5	2,034	2,390	3,068	2,165	2,525	3,205
6	1,977	2,330	3,003	2,093	2,451	3,130
7	1,935	2,285	2,955	2,030	2,395	3,073
8	1,902	2,250	2,917	1,996	2,350	3,024
9	1,875	2,222	2,886	1,961	2,313	2,986
10	1,854	2,198	2,861	1,933	2,283	2,954
11	1,836	2,179	2,839	1,900	2,258	2,926
12	1,821	2,162	2,821	1,889	2,236	2,903
13	1,808	2,148	2,805	1,872	2,218	2,882
14	1,797	2,136	2,790	1,857	2,201	2,864
15	1,788	2,125	2,778	1,844	2,187	2,848
16	1,779	2,115	2,767	1,832	2,174	2,834
17	1,772	2,107	2,757	1,822	2,163	2,821
18	1,765	2,099	2,748	1,812	2,152	2,809
19	1,759	2,092	2,740	1,804	2,143	2,799
20	1,753	2,086	2,733	1,797	2,135	2,789
25	1,732	2,062	2,704	1,767	2,102	2,751
30	1,718	2,046	2,684	1,748	2,080	2,725
35	1,708	2,034	2,670	1,733	2,063	2,706
40	1,700	2,025	2,659	1,723	2,051	2,691
45	1,694	2,018	2,650	1,714	2,041	2,679
50	1,689	2,012	2,643	1,708	2,033	2,669
60	1,682	2,004	2,632	1,697	2,022	2,655
70	1,677	1,998	2,625	1,690	2,013	2,644
80	1,673	1,993	2,619	1,684	2,007	2,636
90	1,670	1,990	2,614	1,680	2,002	2,629
100	1,667	1,987	2,610	1,677	1,998	2,624
200	1,656	1,974	2,594	1,661	1,979	2,601
300	1,653	1,969	2,588	1,656	1,973	2,593
400	1,651	1,967	2,585	1,653	1,970	2,589
500	1,650	1,966	2,582	1,651	1,967	2,585
1000	1,648	1,963	2,579	1,649	1,954	2,581
∞	1,645	1,960	2,576	1,645	1,960	2,576

$a^*(n, \alpha, \beta)$, являющиеся решением уравнения

$$\Phi\left(u_{\frac{1+\alpha}{2}}\frac{1}{\sqrt{n}} + a^*(n, \alpha, \beta)\right) - \Phi\left(u_{\frac{1+\alpha}{2}}\frac{1}{\sqrt{n}} - a^*(n, \alpha, \beta)\right) = \beta$$

приведены в табл. 202.

Задача 306. *Имеется выборка из нормального распределения с известной дисперсией $\sigma^2 = 54$ и неизвестным средним*

$$x_i\colon 1,\ 3,\ 6,\ 14,\ 18,\ 24,\ 31,\ 41,\ 61,\ 63.$$

Необходимо найти односторонние толерантные пределы и двусторонний доверительный интервал при $\alpha = 0{,}95$ и $\beta = 0{,}90$.

Имеем $\bar{x} = 26$. Воспользовавшись табл. 201 и 202, находим значения множителей $a(n,\alpha,\beta) = a(10; 0{,}95; 0{,}90) = 1{,}802$ и $a^*(n,\alpha,\beta) = a^*(10; 0{,}95; 0{,}90) = 1{,}933$.

Тогда односторонние толерантные пределы равны

$$\tau_{\text{в}} = 26 + 1{,}802 \cdot \sqrt{54} = 39{,}24; \quad \tau_{\text{н}} = 26 - 1{,}802 \cdot \sqrt{54} = 12{,}76.$$

Для двустороннего интервала имеем

$$\tau_{\text{в}} = 26 + 1{,}933 \cdot \sqrt{54} = 40{,}20; \quad \tau_{\text{н}} = 26 - 1{,}933 \cdot \sqrt{54} = 11{,}79.$$

Таким образом, можно утверждать с вероятностью $\alpha = 0{,}95$, что в каждом из интервалов $(-\infty; 39{,}24]$; $[12{,}76; \infty)$, $[11{,}79; 40{,}20]$ находится 90% ($\beta = 0{,}90$) всей совокупности значений случайной величины.

4.4.1.2.2. Среднее μ известно, дисперсия σ^2 неизвестна

Односторонние доверительные интервалы находятся по формулам

$$\tau_{\text{в}} = \mu + b(n,\alpha,\beta)s; \quad \tau_{\text{н}} = \mu - b(n,\alpha,\beta)s,$$

где

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2; \quad b(n,\alpha,\beta) = u_\beta\sqrt{\frac{n}{\chi^2_{1-\alpha}(n)}};$$

$\chi^2_{1-\alpha}(n)$ — $(1-\alpha)$-квантиль распределения хи-квадрат с $f = n$ степенями свободы; u_β — β-квантиль стандартного нормального распределения.

Толерантные множители $b(n,\alpha,\beta)$ приведены в табл. 203.

Двусторонний толерантный интервал определятся из соотношения

$$\tau_{\text{в}} = \mu + b^*(n,\alpha,\beta)\, s; \quad \tau_{\text{н}} = \mu - b^*(n,\alpha,\beta)\, s, \quad \text{где} \quad b^*(n,\alpha,\beta) = u_{\frac{1+\beta}{2}}\sqrt{\frac{n}{\chi^2_{1-\alpha}(n)}}.$$

Толерантные множители $b^*(n,\alpha,\beta)$ приведены в табл. 204.

Задача 307. *Для данных задачи* 306 *найти толерантные пределы, исходя из того, что среднее $\mu = 26$ считается известным, а значение дисперсии неизвестно* ($\alpha = 0{,}95$, $\beta = 0{,}90$).

Вычисляем $S^2 = \dfrac{1}{10}\cdot\sum\limits_{i=1}^{10}(x_i - 26)^2 = 490{,}96$ ($S = 21{,}7$). Из табл. 203 и 204 находим толерантные множители $b(n,\alpha,\beta) = b(10, 0{,}95, 0{,}90) = 2{,}042$ для односторонних пределов и $b^*(n,\alpha,\beta) = b^*(10, 0{,}95; 0{,}90) = 2{,}620$ для двустороннего интервала.

Окончательно имеем для односторонних толерантных пределов

$$\tau_{\text{в}} = 26 + 2{,}042 \cdot 21{,}7 = 70{,}31; \quad \tau_{\text{н}} = 26 - 2{,}042 \cdot 21{,}7 = -18{,}31.$$

и для двустороннего интервала

$$\tau_{\text{в}} = 26 + 2{,}620 \cdot 21{,}7 = 82{,}85; \quad \tau_{\text{н}} = 26 - 2{,}620 \cdot 21{,}7 = -30{,}85.$$

Следовательно, можно утверждать, что с вероятностью $\alpha = 0{,}95$ в каждом из интервалов $(-\infty; 70{,}31]$; $[-18{,}31; \infty)$; $[-30{,}85; 82{,}85]$ находится 90% ($\beta = 0{,}90$) всей совокупности значений случайной величины x.

Таблица 203

Множители $b(n, \alpha, \beta)$ для построения односторонних толерантных пределов в случае нормального распределения (μ известно, σ^2 неизвестно) [56]

n	α					
	0,90			0,95		
	β			β		
	0,90	0,95	0,99	0,90	0,95	0,99
2	3,948	6,067	7,167	5,659	7,263	10,272
3	2,904	3,727	5,271	3,742	4,803	6,793
4	2,485	3,190	4,511	3,040	3,902	5,519
5	2,258	2,898	4,099	2,677	3,437	4,860
6	2,114	2,714	3,838	2,455	3,151	4,455
7	2,014	2,586	3,657	2,303	2,956	4,181
8	1,940	2,491	3,522	2,193	2,814	3,980
9	1,883	2,417	3,418	2,108	2,706	3,827
10	1,837	2,358	3,335	2,042	2,620	3,706
11	1,800	2,310	3,267	1,987	2,551	3,607
12	1,768	2,269	3,210	1,942	2,492	3,525
13	1,741	2,235	3,161	1,904	2,443	3,456
14	1,718	2,205	3,119	1,871	2,401	3,396
15	1,698	2,179	3,082	1,842	2,364	3,344
16	1,680	2,156	3,049	1,817	2,332	3,298
17	1,664	2,136	3,020	1,794	2,303	3,257
18	1,650	2,117	2,994	1,774	2,277	3,221
19	1,637	2,101	2,971	1,756	2,254	3,188
20	1,625	2,085	2,949	1,740	2,233	3,158
25	1,579	2,026	2,866	1,676	2,152	3,043
30	1,547	1,985	2,807	1,632	2,095	2,963
35	1,523	1,954	2,764	1,600	2,053	2,904
40	1,504	1,930	2,730	1,574	2,020	2,858
45	1,489	1,911	2,702	1,554	1,994	2,821
50	1,476	1,895	2,680	1,537	1,973	2,790
60	1,456	1,869	2,644	1,511	1,939	2,742
70	1,441	1,850	2,617	1,491	1,913	2,706
80	1,430	1,835	2,595	1,475	1,893	2,678
90	1,420	1,823	2,578	1,462	1,877	2,654
100	1,412	1,812	2,563	1,445	1,863	2,635
200	1,371	1,759	2,488	1,397	1,793	2,536
300	1,353	1,737	2,456	1,374	1,764	2,495
400	1,343	1,724	2,438	1,361	1,747	2,471
500	1,336	1,709	2,417	1,352	1,735	2,455
1000	1,320	1,694	2,395	1,331	1,708	2,415
∞	1,282	1,645	2,326	1,282	1,645	2,326

4.4.1.2.3. Среднее μ и дисперсия σ^2 неизвестны

Односторонние доверительные интервалы находятся по формулам

$$\tau_{\text{в}} = \bar{x} + k\,(n, \alpha, \beta)\,s; \quad \tau_{\text{н}} = \bar{x} - k\,(n, \alpha, \beta)\,s,$$

где $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$; $s^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$; $k(n, \alpha, \beta) = \dfrac{1}{\sqrt{n}}\,t(n-1, u_\beta\sqrt{n}, \alpha)$; $t(m, \delta, q)$ —

Таблица 204

Множители $b^*(n,\alpha,\beta)$ для построения двусторонних толерантных интервалов в случае нормального распределения (μ известно, σ^2 неизвестно) [56]

n	α					
	0,90			0,95		
	β			β		
	0,90	0,95	0,99	0,90	0,95	0,99
2	5,067	6,038	7,936	7,263	8,654	11,373
3	3,727	4,441	5,836	4,803	5,723	7,521
4	3,190	3,801	4,995	3,902	4,650	6,111
5	2,898	3,454	4,539	3,437	4,095	5,382
6	2,714	3,234	4,250	3,151	3,754	4,934
7	2,586	3,081	4,049	2,956	3,522	4,629
8	2,491	2,968	3,900	2,814	3,354	4,407
9	2,417	2,880	3,785	2,706	3,225	4,238
10	2,358	2,810	3,693	2,620	3,122	4,103
11	2,310	2,752	3,617	2,551	3,039	3,994
12	2,269	2,704	3,554	2,492	2,970	3,903
13	2,235	2,663	3,500	2,443	2,911	3,826
14	2,205	2,628	3,453	2,401	2,861	3,760
15	2,179	2,597	3,412	2,364	2,817	3,702
16	2,156	2,569	3,376	2,332	2,778	3,652
17	2,138	2,545	3,344	2,303	2,744	3,607
18	2,117	2,523	3,315	2,277	2,714	3,566
19	2,101	2,503	3,289	2.254	2,686	3,530
20	2,085	2,485	3,266	2,233	2,661	3,497
25	2,026	2,414	3,173	2,152	2,564	3,369
30	1,985	2,365	3,109	2,095	2,496	3,281
35	1,954	2,329	3,060	2,053	2,446	3,215
40	1,930	2,300	3,023	2,020	2,408	3,164
45	1,911	2,277	2,992	1,994	2,376	3,123
50	1,895	2,258	2,967	1,973	2,351	3,089
60	1,869	2,227	2,927	1,939	2,310	3,036
70	1,850	2,205	2,897	1,913	2,280	2,996
80	1,835	2,187	2,874	1,893	2,256	2,965
90	1,823	2,172	2,854	1,877	2,236	2,939
100	1,812	2,160	2,838	1,863	2,220	2,918
200	1,759	2,096	2,755	1,793	2,137	2,808
300	1,737	2,070	2,720	1,764	2,102	2,762
400	1,724	2,054	2,699	1,747	2,082	2,736
500	1,715	2,044	2,686	1,735	2,068	2,718
1000	1,694	2,018	2,652	1,708	2,045	2,675
∞	1,645	1,960	2,326	1,645	1,960	2,326

квантиль нецентрального t-распределения (см. раздел 1.1.23) с $f = m$ степенями свободы и параметром нецентральности δ.

Толерантные множители $k(n,\alpha,\beta)$ приведены в табл. 205.

При $n \geqslant 50$ справедлива аппроксимация [542]

$$k(n,\alpha,\beta) = \frac{2(n-1)}{2(n-1) - u_\alpha^2}\left[u_\beta + \frac{u_\alpha}{\sqrt{2(n-1)}}\sqrt{\frac{2(n-1)}{n} + u_\beta^2 - \frac{u_\beta^2}{n}}\,\right].$$

Таблица 205

Множители $k(n, \alpha, \beta)$ для построения односторонних толерантных пределов в случае нормального распределения (μ и σ^2 неизвестны) [56]

n	α					
	0,90			0,95		
	β			β		
	0,90	0,95	0,99	0,90	0,95	0,99
5	2,745	3,402	4,665	3,413	4,209	5,746
6	2,494	3,093	4,243	3,009	3,709	5,065
7	2,333	2,894	3,971	2,752	3,401	4,644
8	2,220	2,755	3,782	2,582	3,188	4,355
9	2,133	2,650	3,642	2,454	3,032	4,134
10	2,066	2,569	3,533	2,355	2,911	3,981
11	2,012	2,503	3,443	2,276	2,815	3,851
12	1,966	2,448	3,370	2,210	2,736	3,746
13	1,928	2,402	3,298	2,156	2,670	3,658
14	1,896	2,363	3,256	2,109	2,614	3,583
15	1,867	2,329	3,211	2,069	2,566	3,519
16	1,842	2,299	3,172	2,034	2,524	3,456
17	1,820	2,273	3,139	2,002	2,486	3,414
18	1,800	2,249	3,105	1,974	2,458	3,370
19	1,782	2,227	3,077	1,949	2,423	3,330
20	1,766	2,205	3,051	1,926	2,396	3,294
25	1,702	2,132	2,952	1,838	2,292	3,157
30	1,657	2,080	2,883	1,778	2,220	3,063
35	1,624	2,041	2,832	1,733	2,167	2,994
40	1,598	2,010	2,793	1,698	2,126	2,940
45	1,577	1,98 6	2,761	1,669	2,092	2,897
50	1,560	1,965	2,734	1,646	2,065	2,862
60	1,532	1,933	2,693	1,609	2,022	2,807
70	1,512	1,909	2,662	1,582	1,990	2,765
80	1,495	1,890	2,637	1,560	1,964	2,732
90	1,482	1,874	2,617	1,542	1,944	2,706
100	1,471	1,861	2,600	1,527	1,927	2,684
200	1,412	1,793	2,514	1,450	1,832	2,569
300	1,387	1,765	2,477	1,417	1,800	2,521
500	1,362	1,737	2,441	1,386	1,763	2,475
1000	1,338	1,709	2,407	1,355	1,727	2,430
∞	1,282	1,645	2,326	1,282	1,645	2,326

Двусторонний доверительный интервал определяется из соотношения

$$\tau_{\text{в}} = \bar{x} + k^*(n, \alpha, \beta)\, s; \quad \tau_{\text{н}} = \bar{x} - k^*(n, \alpha, \beta)\, s,$$

где $k^*(n, \alpha, \beta)$ — толерантные множители, приведенные в табл. 206.

Для множителей $k^*(n, \alpha, \beta)$ справедливо соотношение

$$k^*(n, \alpha, \beta) = c\sqrt{\frac{n-1}{\chi^2_\alpha(n-1)}},$$

где c — решение уравнения $\Phi\left(\frac{1}{\sqrt{n}} + c\right) - \Phi\left(\frac{1}{\sqrt{n}} - c\right) = \beta$.

Таблица 206

Множители $k^*(n, \alpha, \beta)$ для построения двусторонних толерантных интервалов в случае нормального распределения (μ и σ^2 неизвестны) [56]

n	α					
	0,90			0,95		
	β			β		
	0,90	0,95	0,99	0,90	0,95	0,99
2	15,980	18,800	24,170	32,020	37,670	48,430
3	5,847	6,979	8,974	8,380	9,916	12,860
4	4,166	4,943	6,440	5,369	6,370	8,299
5	3,494	4,152	5,423	4,275	5,079	6,634
6	3,131	3,723	4,870	3,712	4,414	5,775
7	2,902	3,452	4,521	3,369	4,007	5,248
8	2,743	3,264	4,278	3,136	3,732	4,891
9	2,626	3,125	4,098	2,967	3,532	4,631
10	2,535	3,018	3,959	2,839	3,379	4,433
11	2,463	2,933	3,849	2,737	3,259	4,277
12	2,404	2,863	3,758	2,655	3,162	4,150
13	2,355	2,805	3,682	2,587	3,081	4,044
14	2,314	2,756	3,618	2,529	3,012	3,955
15	2,278	2,713	3,562	2,480	2,954	3,878
16	2,246	2,676	3,514	2,437	2,903	3,812
17	2,219	2,643	3,471	2,400	2,858	3,754
18	2,194	2,614	3,433	2,366	2,819	3,702
19	2,172	2,588	3,399	2,337	2,784	3,656
20	2,152	2,564	3,368	2,310	2,752	3,615
25	2,077	2,474	3,251	2,208	2,631	3,457
30	2,025	2,413	3,170	2,140	2,549	3,350
35	1,988	2,368	3,112	2,090	2,490	3,272
40	1,959	2,334	3,066	2,052	2,445	3,213
45	1,935	2,306	3,030	2,021	2,408	3,165
50	1,916	2,284	3,001	1,996	23,379	3,126
60	1,887	2,248	2,955	1,958	2,333	3,066
70	1,865	2,222	2,920	1,929	2,299	3,021
80	1,848	2,202	2,894	1,907	2,720	2,986
90	1,834	2,185	2,872	1,889	2,251	2,958
100	1,822	2,172	2,854	1,874	2,233	2,934
200	1,764	2,102	2,762	1,798	2,143	2,816
300	1,740	2,073	2,725	1,767	2,106	2,767
400	1,726	2,057	2,703	1,749	2,084	2,739
500	1,717	2,046	2,689	1,737	2,070	2,721
1000	1,695	2,019	2,654	1,709	2,036	2,676
∞	1,645	1,960	2,326	1,645	1,960	2,326

При $n \geqslant 50$ может быть использовано приближение Боукера [541]

$$k^*(n, \alpha, \beta) = u_{\frac{1+\beta}{2}} \left(1 + \frac{u_\alpha}{\sqrt{2n}} + \frac{5u_\alpha^2 + 10}{12n}\right).$$

Подробные сведения о толерантных пределах для нормального распределения содержатся в [557, 558], где рассмотрены оценки объема выборки n по заданным β и α, оценки толерантных интервалов для нескольких выборок с равными дисперсиями [558].

Задача 308. *Для ряда значений* ($n = 20$)

$$x_i\colon 2,\ 4,\ 5,\ 7,\ 8,\ 11,\ 13,\ 17,\ 20,\ 24,\ 23,\ 25,\ 18,\ 18,\ 17,\ 17,\ 14,\ 11,\ 8,\ 1$$

найти толерантные пределы при $\alpha = 0{,}95$ *и* $\beta = 0{,}95$.

Вычисляем $\bar{x} = \frac{1}{20} \cdot \sum\limits_{i=1}^{20} x_i = 13{,}15$; $s^2 = \frac{1}{20-1} \cdot \sum\limits_{i=1}^{20} (x_i - 13{,}15)^2 = 53{,}50$ ($s = 7{,}314$).

Из табл. 205 и 206 для $n = 20$, $\alpha = 0{,}95$ и $\beta = 0{,}95$ находим толерантные коэффициенты $k(n, \alpha, \beta) = 2{,}396$ и $k^*(n,\alpha,\beta) = 2{,}752$. Тогда имеем для односторонних толерантных пределов:

$$\tau_{\text{в}} = \bar{x} + k\,(n, \alpha, \beta) \cdot s = 13{,}15 + 2{,}396 \cdot 7{,}314 = 30{,}67;$$
$$\tau_{\text{н}} = \bar{x} - k\,(n, \alpha, \beta) \cdot s = 13{,}15 - 2{,}396 \cdot 7{,}314 = -4{,}374.$$

Для двустороннего толерантного интервала:

$$\tau_{\text{в}} = \bar{x} + k^*(n, \alpha, \beta) \cdot s = 13{,}15 + 2{,}752 \cdot 7{,}314 = 33{,}28;$$
$$\tau_{\text{н}} = \bar{x} - k^*(n, \alpha, \beta) \cdot s = 13{,}15 - 2{,}752 \cdot 7{,}314 = -6{,}98.$$

Следовательно, с вероятностью $\alpha = 0{,}95$ в каждом из интервалов $(-\infty; 30{,}67]$; $[-4{,}37; \infty)$; $[-6{,}98; 33{,}28]$ будет находиться 95% ($\beta = 0{,}95$) всей совокупности значений случайной величины.

4.4.1.2.4. Толерантные пределы, основанные на выборочном размахе

Для быстрой оценки толерантных пределов без предварительной выборочной оценки стандартного отклонения s можно использовать выборочный размах $\omega_n = x_n - x_1$ или среднее нескольких выборочных размахов, вычисляемых для m подвыборок равного объема $n' = \frac{n}{m}$ [543–545], $\omega_{mn} = \frac{1}{m} \sum\limits_{i=1}^{m} \omega_i$. Границы двусторон-

Таблица 207

Множители $d(1, n, \alpha, \beta)$ для определения двустороннего толерантного интервала с использованием размахов (при $m = 1$, без разбиения на подвыборки) [543]

n	α					
	0,90			0,95		
	β			β		
	0,90	0,95	0,99	0,90	0,95	0,99
2	10,381	12,427	16,262	20,811	24,908	32,593
3	2,860	3,440	4,534	4,109	4,939	6,505
4	1,753	2,118	2,806	2,270	2,738	3,642
5	1,331	1,613	2,147	1,639	1,983	2,635
6	1,108	1,347	1,800	1,325	1,607	2,143
7	0,970	1,182	1,583	1,137	1,382	1,848
8	0,875	1,069	1,435	1,011	1,232	1,650
9	0,805	0,985	1,326	0,920	1,123	1,509
10	0,752	0,922	1,243	0,852	1,042	1,401
11	0,709	0,871	1,177	0,798	0,977	1,317
12	0,674	0,829	1,122	0,755	0,925	1,249
13	0,645	0,795	1,077	0,719	0,882	1,193
14	0,621	0,765	1,038	0,688	0,846	1,145
15	0,600	0,740	1,005	0,662	0,815	1,104

Таблица 208

Множители $d(m, n', \alpha, \beta)$ для определения двустороннего толерантного интервала с использованием размахов [543]

m	α					
	0,90			0,95		
	β			β		
	0,90	0,95	0,99	0,90	0,95	0,99
$n' = 5$						
2	1,008	1,237	1,669	1,154	1,412	1,901
3	0,897	1,107	1,505	0,997	1,227	1,664
4	0,838	1,039	1,418	0,917	1,133	1,543
5	0,801	0,996	1,363	0,867	1,075	1,468
6	0,775	0,965	1,325	0,832	1,034	1,416
7	0,755	0,942	1,296	0,806	1,004	1,377
8	0,740	0,924	1,273	0,786	0,980	1,347
9	0,727	0,910	1,255	0,770	0,961	1,323
10	0,717	0,898	1,240	0,756	0,946	1,303
11	0,708	0,888	1,227	0,745	0,933	1,286
12	0,700	0,879	1,216	0,736	0,921	1,272
13	0,694	0,871	1,206	0,727	0,911	1,260
14	0,688	0,865	1,198	0,720	0,903	1,248
15	0,683	0,859	1,190	0,713	0,895	1,239
$n' = 10$						
2	0,631	0,782	1,067	0,689	0,852	1,160
3	0,584	0,727	0,998	0,626	0,778	1,066
4	0,558	0,697	0,960	0,592	0,738	1,015
5	0,541	0,677	0,935	0,570	0,713	0,982
6	0,528	0,663	0,917	0,555	0,694	0,959
7	0,519	0,652	0,903	0,543	0,681	0,941
8	0,512	0,644	0,893	0,533	0,670	0,927
9	0,506	0,637	0,884	0,526	0,661	0,916
10	0,501	0,631	0,877	0,520	0,654	0,907
11	0,496	0,626	0,871	0,514	0,647	0,899
12	0,493	0,622	0,865	0,510	0,642	0,892
13	0,489	0,618	0,861	0,506	0,637	0,886
14	0,486	0,615	0,856	0,502	0,633	0,881
15	0,484	0,612	0,853	0,499	0,630	0,876

него толерантного интервала имеют вид

$$\tau_{\text{в}} = \bar{x} + d(m, n', \alpha, \beta)\omega_{mn}; \quad \tau_{\text{н}} = \bar{x} - d(m, n', \alpha, \beta)\omega_{mn},$$

где $d(m, n', \alpha, \beta)$ — толерантные коэффициенты, значения которых приведены в табл. 207 (при отсутствии разбиения на подвыборки, т. е. когда $m = 1$) и в табл. 208 (при $m > 1$).

Задача 309. *Для ряда значений задачи 308 построить двусторонний толерантный интервал с помощью размахов.*

Учитывая, что при $n \geqslant 20$ существенно уменьшается эффективность оценки s с помощью размаха, разобьем выборку на две подвыборки ($m = 2$) равного объема $n' = 10$ ($n = n'$; $m = 2$). Для $\alpha = 0{,}95$ и $\beta = 0{,}95$, воспользовавшись табл. 208, находим $d(2; 10; 0{,}95; 0{,}95) = 0{,}946$.

Для подвыборок имеем размахи $\omega_1 = 24 - 2 = 22$, $\omega_2 = 25 - 1 = 24$ и средний размах $\omega_{mn} = \dfrac{22+24}{2} = 23$. Далее вычисляем

$$\tau_в = 13{,}15 + 0{,}946 \cdot 23 = 34{,}908; \quad \tau_н = 13{,}15 - 0{,}946 \cdot 23 = -8{,}608.$$

Следовательно, с вероятностью $\alpha = 0{,}95$ в интервале $[-8{,}61; 34{,}91]$ будет находиться 95% ($\beta = 0{,}95$) всей совокупности значений случайной величины.

Отсюда видно, что толерантный интервал близок к интервалу, полученному в задаче 308, что свидетельствует о достаточной эффективности использования размахов.

4.4.1.2.5. Толерантные пределы для выборочных дисперсий

Будем искать интервал, который с вероятностью α включает β-ю часть совокупности выборочных дисперсий

$$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2, \quad \text{где} \quad \bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}.$$

Очевидно, что β-квантиль распределения s_i^2 есть

$$q = \sigma^2 \frac{\chi_\beta^2(n-1)}{n-1},$$

так как $\dfrac{(n-1)\,s_i^2}{\sigma^2}$ имеет распределение хи-квадрат с $f = n - 1$ степенями свободы. Для оценки σ^2 используем среднюю оценку по совокупности значений $s_1^2, s_2^2, \ldots, s_m^2$. Тогда оценкой σ^2 будет величина

$$s^2 = \frac{(n-1)(s_1^2 + s + \ldots + s_m^2)}{m(n-1)} = \frac{1}{m}\sum_{i=1}^{m} s_i^2.$$

Оценке s^2 соответствует $f = m(n-1)$ степеней свободы. Тогда толерантный интервал, включающий в себя β-ю часть всей совокупности значений $s_1^2, s_2^2, \ldots, s_m^2$ (каждая оценка s_i^2 получена по выборке равного объема n) будет равен [546]

$$\tau_в = \frac{m\chi_\beta^2(n-1)}{\chi_{1-\alpha}^2[m\,(n-1)]}s^2; \quad \tau_н = 0.$$

Для стандартного отклонения $s = \sqrt{s^2}$ толерантный интервал определятся соотношениями [546] $\tau_в = \left\{\dfrac{m\chi_\beta^2(n-1)}{\chi_{1-\alpha}^2[m\,(n-1)]}\right\}^{\frac{1}{2}}$; $\tau_н = 0$.

Напомним, что при $f > 30$ справедлива аппроксимация $\chi_\beta^2 = \dfrac{1}{2}\left(\sqrt{2f-1} + u_\beta\right)^2$.

Задача 310. *В результате наблюдений над $m = 10$ выборками из нормального распределения получены следующие выборочные значения дисперсий:*

$$s_i^2\colon\ 25,\ 41,\ 48,\ 37,\ 21,\ 64,\ 91,\ 49,\ 59,\ 78.$$

Оценки s_i^2 получены на выборках равного объема $n = 20$. Необходимо найти толерантный интервал для значений s_i^2 при $\alpha = 0{,}95$ и $\beta = 0{,}90$.

Имеем $s^2 = \dfrac{1}{10}\cdot\sum\limits_{i=1}^{10} s_i^2 = 51{,}3$. По табл. 55 χ^2-распределения (или используем аппроксимации из раздела 1.1.8):

$$\chi_\beta^2(n-1) = \chi_{0,90}^2(19) = 27{,}202; \quad \chi_{1-\alpha}^2[m\cdot(n-1)] = \chi_{0,05}^2(190) = 159{,}11.$$

Вычисляем $\tau_{\text{в}} = \dfrac{10 \cdot 0{,}946 \cdot 23}{159{,}11} \cdot 51{,}3 = 87{,}71$. Следовательно, можно утверждать, что с вероятностью $\alpha = 0{,}95$ в интервале $[0; 87{,}71]$ находится 90% ($\beta = 0{,}9$) всей совокупности значений s_i^2. Соответствующий интервал для стандартного отклонения s будет $[0; \sqrt{87{,}71}] = [0; 9{,}36]$.

4.4.2. Непараметрические толерантные пределы

Иногда представляет интерес при неизвестном непрерывном распределении определить тот минимальный объем выборки $n_{\min}$, при котором можно утверждать, что с вероятностью α для любой непрерывной совокупности между минимальным ($x_{\min}$) и максимальным ($x_{\max}$) значениями выборки, извлеченной из нее, заключена β-я доля всей совокупности. В этом случае мы говорим о двустороннем непараметрическом толерантном интервале.

Если через $n^*_{\min}$ обозначить такой минимальный объем выборки, при котором можно утверждать, что с вероятностью α в каждом из интервалов $(-\infty; x_{\max})$, $(x_{\min}, \infty)$ содержится β-я доля любой непрерывной совокупности, то говорят о непараметрических односторонних толерантных интервалах. Между $n_{\min}$, α и β для двустороннего толерантного интервала справедливо соотношение Уилкса [547]

$$n_{\min}\beta^{n_{\min}-1} - (n_{\min} - 1)\,\beta^{n_{\min}} \leqslant 1 - \alpha.$$

Для односторонних интервалов справедливо неравенство $\beta^{n^*_{\min}} \leqslant 1-\alpha$. Значения $n_{\min}$ и $n^*_{\min}$, удовлетворяющие приведенным неравенствам приведены в табл. 209.

Таблица 209

Минимальные объемы выборок для построения непараметрических (двустороннего, $n_{\min}$ — верхняя строка **и одностороннего, $n^*_{\min}$** — нижняя строка) **толерантных интервалов, основанных на наибольших и наименьших выборочных значениях** [56]

α	β						
	0,50	0,70	0,75	0,80	0,90	0,95	0,99
0,90	7	12	15	18	38	77	388
	4	7	9	11	22	45	30
0,95	8	14	18	22	46	93	473
	5	9	11	14	29	59	299
0,99	11	20	33	31	64	130	662
	7	3	17	21	44	90	459

Практический интерес представляет и обратная задача: для заданных значений n и α определить наименьшую долю β совокупности, относительно которой можно утверждать, что с вероятностью α эта доля заключена между крайними членами выборки объема n, либо превосходит наименьшее значение выборки ($x_{\min}$), либо меньше наибольшего значения выборки ($x_{\max}$).

Для двусторонних интервалов при $n \geqslant 50$ справедливо соотношение $\beta = 1 - \dfrac{k}{n}$, где k является решением уравнения

$$1 + k - (1 - \alpha)e^k = 0.$$

Некоторые решения этого уравнения для наиболее употребительных значений α таковы:

α	0,9	0,95	0,99
k	3,8897	4,7439	6,6384

Для односторонних интервалов справедливо соотношение

$$\beta = 1 + \frac{\ln(1-\alpha)}{n}.$$

В табл. 210 и 211 приведены значения β, соответствующие различным α и n, соответственно для двустороннего и одностороннего толерантных интервалов.

Таблица 210

Наименьшая доля β совокупности, заключенная внутри непараметрического двустороннего толерантного интервала (между $x_{\min}$ и $x_{\max}$ в выборке объема n) [56]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
2	0,0513	0,0253	0,0050	34	0,8903	0,8679	0,8202
3	0,1958	0,1354	0,0589	35	0,8934	0,8715	0,8249
4	0,3205	0,2486	0,1409	36	0,8962	0,8749	0,8294
5	0,4101	0,3426	0,2221	37	0,8989	0,8781	0,8337
6	0,4897	0,4182	0,2943	38	0,9015	0,8811	0,8377
7	0,5474	0,4793	0,3566	39	0,9039	0,8840	0,8416
8	0,5938	0,5293	0,4101	40	0,9062	0,8868	0,8453
9	0,6316	0,5709	0,4560	41	0,9084	0,8894	0,8488
10	0,6632	0,6058	0,4956	42	0,9105	0,8920	0,8521
11	0,6898	0,6356	0,5302	43	0,9125	0,8944	0,8554
12	0,7125	0,6613	0,5605	44	0,9145	0,8967	0,8584
13	0,7322	0,6873	0,5872	45	0,9163	0,8989	0,8614
14	0,7493	0,7033	0,6109	46	0,9181	0,9010	0,8642
15	0,7644	0,7206	0,6321	47	0,9197	0,9030	0,8669
16	0,7778	0,7360	0,6512	48	0,9214	0,9049	0,8695
17	0,7898	0,7499	0,6684	49	0,9229	0,9068	0,8721
18	0,8005	0,7623	0,6840	50	0,9244	0,9086	0,8745
19	0,8102	0,7736	0,6982	60	0,9367	0,9234	0,8944
20	0,8190	0,7839	0,7112	70	0,9456	0,9340	0,9089
21	0,8271	0,7933	0,7232	80	0,9952	0,9421	0,9199
22	0,8344	0,8019	0,7342	90	0,9575	0,9484	0,9285
23	0,8412	0,8098	0,7443	100	0,9617	0,9534	0,9355
24	0,8474	0,8171	0,7538	200	0,9807	0,9765	0,9673
25	0,8531	0,8239	0,7625	300	0,9871	0,9843	0,9781
26	0,8585	0,8302	0,7707	400	0,9903	0,9882	0,9835
27	0,8634	0,8360	0,7783	500	0,9922	0,9905	0,9868
28	0,8681	0,8415	0,7854	600	0,9935	0,9921	0,9890
29	0,8724	0,8466	0,7923	700	0,9945	0,9932	0,9906
30	0,8764	0,8514	0,8004	800	0,9951	0,9941	0,9917
31	0,8802	0,8559	0,8043	900	0,9957	0,9947	0,9926
32	0,8838	0,8602	0,8099	1000	0,9961	0,9953	0,9934
33	0,8872	0,8641	0,9152	5000	0,9992	0,9991	0,9987

Таблица 211

Наименьшая доля совокупности β, заключенная внутри непараметрического одностороннего толерантного интервала (больше $x_{\min}$ или меньше $x_{\max}$ в выборке объема n) [56]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
2	0,3162	0,2236	0,1000	33	0,9326	0,9132	0,8697
3	0,4642	0,3684	0,2154	34	0,9345	0,9157	0,8733
4	0,5623	0,4729	0,3162	35	0,9363	0,9180	0,8767
5	0,6310	0,5493	0,3981	36	0,9380	0,9202	0,8799
6	0,6813	0,6070	0,4642	37	0,9397	0,9222	0,8830
7	0,7197	0,6518	0,5179	38	0,9412	0,9242	0,8859
8	0,7499	0,6877	0,5623	39	0,9427	0,9261	0,8886
9	0,7743	0,7169	0,5995	40	0,9441	0,9278	0,8913
10	0,7943	0,7411	0,6310	41	0,9454	0,9295	0,8938
11	0,8110	0,7791	0,6813	42	0,9467	0,9312	0,8952
12	0,8254	0,7791	0,6813	43	0,9479	0,9327	0,8984
13	0,8377	0,7942	0,7017	44	0,9490	0,9342	0,9006
14	0,8483	0,8074	0,7197	45	0,9501	0,9356	0,9027
15	0,8577	0,8190	0,7356	46	0,9512	0,9370	0,9047
16	0,8660	0,8293	0,7499	47	0,9522	0,9382	0,9067
17	0,8733	0,8384	0,7627	48	0,9532	0,9395	0,9085
18	0,8799	0,8467	0,7743	49	0,9541	0,9407	0,9103
19	0,8859	0,8541	0,7848	50	0,9550	0,9418	0,9120
20	0,8913	0,8609	0,7943	60	0,9624	0,9513	0,9261
21	0,8962	0,8671	0,8031	70	0,9676	0,9581	0,9363
22	0,9006	0,8727	0,8111	80	0,9716	0,9632	0,9441
23	0,9047	0,8779	0,8185	90	0,9747	0,9673	0,9501
24	0,9085	0,8827	0,8254	100	0,9772	0,9705	0,9550
25	0,9120	0,9871	0,8318	200	0,9886	0,9851	0,9772
26	0,9152	0,8912	0,8377	300	0,9924	0,9901	0,9848
27	0,9183	0,8950	0,8432	400	0,9943	0,9925	0,9886
28	0,9211	0,8985	0,8483	500	0,9954	0,9940	0,9908
29	0,9237	0,9010	0,8532	700	0,9967	0,9957	0,9934
30	0,9261	0,9050	0,8577	800	0,9971	0,9963	0,9943
31	0,9284	0,9079	0,8620	900	0,9974	0,9967	0,9949
32	0,9306	0,9106	0,8660	1000	0,9977	0,9970	0,9954

Задача 311. *Определить объем выборки $x_1, x_2, \ldots, x_n$, для которой с вероятностью $\alpha = 0{,}95$ между $x_{\min}$ и $x_{\max}$ будет заключаться не менее $\beta = 0{,}99$ всей совокупности значений x.*

Из табл. 209 для $\alpha = 0{,}95$ и $\beta = 0{,}99$ находим $n_{\min} = 473$, т.е. в выборке объема $n = 473$ между $x_{\min}$ и $x_{\max}$ будет располагаться не менее 99% всех значений x_i, извлеченных из любой непрерывной совокупности.

Задача 312. *Определить минимальный объем выборки в условиях задачи* 311 *для одностороннего толерантного интервала.*

Из табл. 209 находим $n_{\min} = 299$. Следовательно, для выборки объема $n = 299$ можно утверждать, что левее $x_{\max}$ либо правее $x_{\min}$ будет находиться 99% всей совокупности.

Задача 313. *Определить долю членов непрерывной совокупности, которая с вероятностью $\alpha = 0{,}90$ будет заключена между $x_{\min}$ и $x_{\max}$ в выборке объема $n = 49$.*

Из табл. 210 для $n = 49$ и $\alpha = 0{,}90$ находим $\beta = 0{,}9229$. Следовательно, в выборке объема $n = 49$ между $x_{\min}$ и $x_{\max}$ будет находиться не менее 92,29% всей совокупности случайных величин.

По аналогии для одностороннего интервала из табл. 211 получаем для $\alpha = 0{,}90$ и $n = 49$: $\beta = 0{,}9541$, т. е. в выборке объема $n = 49$ справа от $x_{\min}$ либо слева от $x_{\max}$ будет располагаться 95,41% всей совокупности.

4.4.3. Толерантные пределы для будущих наблюдений и прогнозирование

4.4.3.1. Прогнозные интервалы Холла–Прейри

Иногда в практических задачах анализа надежности и долговечности технических систем требуется определить границы интервалов, аналогичных толерантным, но имеющих смысл „предсказывающих“.

Сформулируем следующую задачу. Имеется выборка объема n из нормально распределенной совокупности, по которой определены выборочные среднее $\bar{x}$ и стандартное отклонение s. Необходимо найти границу интервала $[\bar{x} - r(m, n, k, \alpha); \infty)$ для которого справедливо утверждение, что с вероятностью α не менее m из k последующих наблюдений будут превышать значение $\bar{x} - r(m, n, k, \alpha)s$.

Эта задача рассмотрена в [548]. Коэффициенты оценок $r(m, n, k, \alpha)$, рассчитанные авторами работы, приведены в табл. 212.

Аналогичная задача рассмотрена в упомянутой работе [548] и для экспоненциального распределения. В этом случае искомый интервал определяется как $[b(m, n, k, \alpha)\sum_{i=1}^{n} x_i; \infty)$. Коэффициенты $b(m, n, k, \alpha)$ приведены в табл. 213. Табл. 213 применима и для решения задачи при распределении Вейбулла с известным параметром β (см. раздел 1.1.5). В этом случае искомый интервал будет равен

$$\left\{\left[b\,(m, n, k, \alpha)\sum_{i=1}^{n} x_i^{\beta}\right]^{\frac{1}{\beta}};\infty\right\}.$$

Задача 314. *В результате испытаний $n = 10$ приборов получены следующие результаты x_i: 12, 13, 17, 21, 29, 35, 46, 57, 61, 70. Необходимо определить значение $x(0{,}95)$, которое с вероятностью $\alpha = 0{,}95$ будет превзойдено не менее, чем $m = 4$ приборами из $k = 8$ испытанных дополнительно. Задачу решить, исходя из предположений нормальности и экспоненциальности распределения случайной величины x.*

Рассмотрим сначала случай нормального распределения.

Вычисляем по выборке $\sum_{i=1}^{n} x_i = 361$; $s = 21{,}247$. Из табл. 212 для $m = 4$, $n = 10$, $k = 8$ и $\alpha = 0{,}95$ находим коэффициент $r(4; 10; 8; 0{,}95) = 1{,}3$.

Следовательно, искомое значение равно $x_{0,95} = \bar{x} - rs = 36{,}1 - 1{,}3 \cdot 21{,}747 = 8{,}479$, т. е. с достоверностью $\alpha = 0{,}95$ из 8 испытанных прибора 4 будут иметь параметр $x > 8{,}479$.

В случае экспоненциального распределения имеем $\sum_{i=1}^{n} x_i = 361$, находим из табл. 213 коэффициент $b(4; 10; 8; 0{,}95) = 0{,}030$. Тогда искомая величина равна $x(0{,}95) = b \times \times \sum_{i=1}^{n} x_i = 0{,}030 \cdot 361 = 10{,}83$. Следовательно, с вероятностью $\alpha = 0{,}95$ у 4 приборов из будущих 8 испытываемых значение параметра $x(0{,}95)$ будет превышать 10,83.

Таблица 212

Коэффициенты $r(m, n, k, \alpha)$, для которых вероятность того, что в m или более из будущих k наблюдений нормального распределения величина x будет превышать $\bar{x} - r(m, n, k, \alpha)s$, равна α [548]

m	n	k	r	m	n	k	r	m	n	k	r	m	n	k	r
$\alpha = 0{,}90$															
5	2	2	2	10	2	2	2	15	2	2	1,8	20	2	2	1,7
5	2	3	1	10	2	3	1	15	2	3	1,1	20	2	3	1,2
5	2	4	0,7	10	2	4	0,6	15	2	4	0,5	20	2	4	0,6
5	2	5	0,4	10	2	5	0,25	15	2	5	0,3	20	2	5	0,4
5	2	6	0,2	10	2	6	0	15	2	6	0,15	20	2	6	0,2
5	2	8	−0,1	10	2	8	−0,2	15	2	8	0	20	2	8	−0,1
5	2	10	−0,25	10	2	10	−0,3	15	2	10	−0,2	20	2	10	−0,3
5	2	12	−0,3	10	2	12	−0,5	15	2	12	−0,4	20	2	12	−0,5
5	4	4	2,8	10	4	4	2,3	15	4	4	2,5	20	4	4	2,1
5	4	5	2,1	10	4	5	1,5	15	4	5	1,5	20	4	5	1,3
5	4	6	1,5	10	4	6	1,2	15	4	6	1	20	4	6	1
5	4	8	0,75	10	4	8	0,6	15	4	8	0,55	20	4	8	0,5
5	4	10	0,5	10	4	10	0,3	15	4	10	0,3	20	4	10	0,2
5	4	12	0,25	10	4	12	0,2	15	4	12	0,05	20	4	12	0
5	6	8	1,6	10	6	6	2,5	15	6	8	1,25	20	6	6	2,3
5	6	10	1	10	6	8	1,5	15	6	10	0,75	20	6	8	1,25
5	6	12	0,75	10	6	10	1	15	6	12	0,5	20	6	10	0,8
5	6	16	0,4	10	6	12	0,6	15	6	16	0,2	20	6	12	0,5
5	8	8	3,4	10	6	16	0,25	15	8	8	2,5	20	6	16	0,15
5	8	10	1,9	10	8	8	2,7	15	8	10	1,4	20	8	8	2,5
5	8	12	1,3	10	8	10	1,5	15	8	12	1	20	8	10	1,5
5	8	16	0,8	10	8	12	1	15	8	16	0,45	20	8	12	1
5	10	10	3,5	10	8	16	0,5	15	10	10	2,5	20	8	16	0,45
5	10	12	2	10	10	10	2,8	15	10	12	1,5	20	10	10	2,5
5	10	16	1,2	10	10	12	1,5	15	10	16	0,8	20	10	12	1,5
5	10	20	0,75	10	10	16	0,9	15	10	20	0,5	20	10	16	0,7
5	12	12	3,75	10	10	20	0,5	15	12	12	2,5	20	12	12	0,5
5	12	16	1,75	10	12	12	2,9	15	12	16	1,2	20	12	16	1,2
5	12	20	1,1	10	12	16	1,3	15	12	20	0,7	20	12	20	0,7
5	14	16	2,3	10	12	20	0,75	15	14	16	1,6	20	14	16	1,7
5	14	20	1,5	10	14	16	1,8	15	14	20	1	20	14	20	1
5	16	16	3,5	10	14	20	1,1	15	16	16	2,3	20	16	16	2,5
5	16	20	2	10	16	16	3	15	16	20	1,3	20	16	20	1,4
5	16	25	1,25	10	16	20	1,5	15	16	25	0,7	20	16	25	0,8
5	18	20	2,1	10	16	25	0,95	15	18	20	1,7	20	18	20	1,8
5	18	25	1,5	10	18	20	2	15	18	25	1	20	18	25	1
5	20	20	4	10	18	25	1,3	15	20	20	2,5	20	20	20	2,5
5	20	25	2	10	20	20	3,2	15	20	25	1,25	20	20	25	1,4
5	22	25	2,5	10	20	25	1,5	15	22	25	1,6	20	22	25	1,7
5	24	25	3,5	10	22	25	1,9	15	24	25	2,2	20	24	25	2,3
$\alpha = 0{,}95$															
5	2	2	3	5	12	20	1	15	6	16	0,3	20	4	4	2,5
5	2	3	2,5	5	14	16	2	15	8	8	2,8	20	4	5	1,7
5	2	4	1,5	5	14	20	1,3	15	8	10	1,7	20	4	6	1,2
5	2	5	1	5	16	12	3	15	8	12	1,2	20	4	8	0,7
5	2	6	0,6	5	16	16	1,7	15	8	16	0,6	20	4	10	0,4
5	2	8	0,4	5	16	20	1,15	15	10	10	2	20	4	12	0,2

Окончание таблицы 212

m	n	k	r	m	n	k	r	m	n	k	r	m	n	k	r
$\alpha = 0{,}95$															
5	2	10	0,1	5	18	16	2,1	15	10	12	1,8	20	6	6	2,7
5	2	12	−0,1	5	18	20	1,1	15	10	16	1	20	6	8	1,5
5	4	4	3,5	5	20	16	3,7	15	12	12	0,6	20	6	10	1
5	4	5	2,5	5	20	20	1,7	15	12	16	3,1	20	6	12	0,7
5	4	6	2	5	24	25	2,7	15	12	20	1,5	20	6	16	0,3
5	4	8	1,3	15	2	2	2,5	15	14	16	2	20	8	8	3
5	4	10	0,8	15	2	3	1,5	15	14	20	1,25	20	8	10	1,5
5	4	12	0,5	15	2	4	0,9	15	16	16	3,2	20	8	12	1,1
5	6	6	4	15	2	5	0,5	15	16	20	1	20	8	16	0,6
5	6	8	2,5	15	2	6	0,3	15	16	25	1	20	10	10	3
5	6	10	1,7	15	2	8	0,05	15	18	20	2	20	10	12	1,7
5	6	12	1,2	15	2	10	−0,1	15	18	25	1,2	20	10	16	1
5	6	16	0,6	15	2	12	−0,4	15	20	20	3,4	20	10	20	0,6
5	8	8	4,5	15	4	4	2,5	15	20	25	1,55	20	14	16	1,9
5	8	10	2,7	15	4	5	2	15	22	25	2	20	14	20	1,15
5	8	12	1,8	15	4	6	1,3	20	2	2	2,5	20	16	16	3,4
5	8	16	1,1	15	4	8	0,75	20	2	3	1,7	20	16	20	1,5
5	10	10	3,5	15	4	10	0,5	20	2	4	1	20	16	25	1
5	10	12	2	15	4	12	0,2	20	2	5	0,6	20	18	20	2
5	10	16	1	15	6	6	2,8	20	2	6	0,5	20	18	25	1,3
5	10	20	0,7	15	6	8	1,5	20	2	8	0,2	20	20	20	3,2
5	12	12	2,5	15	6	10	1	20	2	10	0	20	20	25	1,5
5	12	16	1,5	15	6	12	0,6	20	2	12	−0,2	20	22	25	2

Таблица 213

Коэффициенты $b(m,n,k,\alpha)$, для которых вероятность того, что в m или более из будущих k наблюдений экспоненциального распределения величина x будет превышать $b(m,n,k,\alpha)\sum_{i=1}^{n} x_i$, равна α [548]

m	n	k	b	m	n	k	b	m	n	k	b	m	n	k	b
$\alpha = 0{,}90$															
5	2	1	0,015	10	2	2	0,004	15	2	2	0,004	20	2	3	0,010
5	2	2	0,035	10	2	3	0,021	15	2	3	0,012	20	2	4	0,020
5	2	3	0,070	10	2	4	0,040	15	2	4	0,022	20	2	5	0,026
5	2	4	0,100	10	2	5	0,050	15	2	5	0,032	20	2	6	0,030
5	2	5	0,15	10	2	6	0,070	15	2	6	0,050	20	2	8	0,045
5	2	6	0,18	10	2	8	0,090	15	2	8	0,060	20	2	10	0,052
5	2	8	0,20	10	2	10	0,100	15	2	10	0,070	20	2	12	0,060
5	2	10	0,22	10	2	12	0,111	15	2	12	0,080	20	2	16	0,070
5	2	12	0,26	10	2	16	0,114	15	2	16	0,100	20	2	20	0,080
5	2	16	0,28	10	2	20	0,116	15	2	20	0,110	20	2	25	0,090
5	2	20	0,32	10	2	25	0,118	15	2	25	0,130	20	4	4	0,0015
5	4	3	0,006	10	4	4	0,003	15	4	4	0,0015	20	4	5	0,005
5	4	4	0,015	10	4	5	0,012	15	4	5	0,007	20	4	6	0,009
5	4	5	0,040	10	4	6	0,020	15	4	6	0,017	20	4	8	0,020
5	4	6	0,080	10	4	8	0,040	15	4	8	0,028	20	4	10	0,028
5	4	8	0,100	10	4	10	0,055	15	4	10	0,038	20	4	12	0,035

Продолжение таблицы 213

m	n	k	b	m	n	k	b	m	n	k	b	m	n	k	b
$\alpha = 0{,}90$															
5	4	10	0,13	10	4	12	0,070	15	4	12	0,046	20	4	16	0,045
5	4	12	0,18	10	4	16	0,090	15	4	16	0,065	20	4	20	0,058
5	4	16	0,20	10	4	20	0,100	15	4	20	0,075	20	4	25	0,068
5	4	20	0,23	10	4	25	0,120	15	4	25	0,090	20	6	6	0,0009
5	6	5	0,004	10	6	6	0,002	15	6	6	0,001	20	6	8	0,009
5	6	6	0,030	10	6	8	0,016	15	6	8	0,011	20	6	10	0,015
5	6	8	0,065	10	6	10	0,030	15	6	10	0,020	20	6	12	0,020
5	6	10	0,080	10	6	12	0,040	15	6	12	0,028	20	6	16	0,030
5	6	12	0,120	10	6	16	0,060	15	6	16	0,044	20	6	20	0,040
5	6	16	0,150	10	6	20	0,075	15	6	20	0,054	20	6	25	0,050
5	6	20	0,180	10	6	25	0,095	15	6	25	0,068	20	8	8	0,0007
5	8	6	0,003	10	8	8	0,0015	15	8	8	0,001	20	8	10	0,005
5	8	8	0,021	10	8	10	0,012	15	8	10	0,009	20	8	12	0,010
5	8	10	0,045	10	8	12	0,021	15	8	12	0,017	20	8	16	0,021
5	8	12	0,075	10	8	16	0,040	15	8	16	0,030	20	8	20	0,030
5	8	16	0,110	10	8	20	0,055	15	8	20	0,040	20	8	25	0,040
5	8	20	0,150	10	8	25	0,071	15	8	25	0,052	20	10	10	0,0005
5	10	8	0,002	10	10	10	0,001	15	10	10	0,001	20	10	12	0,004
5	10	10	0,018	10	10	12	0,009	15	10	12	0,0065	20	10	16	0,016
5	10	12	0,050	10	10	16	0,028	15	10	16	0,020	20	10	20	0,022
5	10	16	0,080	10	10	20	0,040	15	10	20	0,030	20	10	25	0,030
5	10	20	0,110	10	10	25	0,058	15	10	25	0,040	20	12	12	0,0004
5	12	10	0,001	10	12	12	0,001	15	12	12	0,001	20	12	16	0,008
5	12	12	0,030	10	12	16	0,016	15	12	16	0,012	20	12	20	0,017
5	12	16	0,056	10	12	20	0,030	15	12	20	0,021	20	12	25	0,024
5	12	20	0,085	10	12	25	0,045	15	12	25	0,033	20	14	16	0,003
5	14	12	0,011	10	14	16	0,0065	15	14	16	0,005	20	14	20	0,011
5	14	16	0,040	10	14	20	0,020	15	14	20	0,015	20	14	25	0,019
5	14	20	0,068	10	14	25	0,035	15	14	25	0,027	20	16	16	0,0002
5	16	12	0,015	10	16	16	0,001	15	16	20	0,009	20	16	20	0,006
5	16	16	0,024	10	16	20	0,012	15	16	25	0,020	20	16	25	0,015
5	16	20	0,050	10	16	25	0,026	15	18	20	0,037	20	18	20	0,0025
5	18	16	0,010	10	18	20	0,005	15	18	25	0,017	20	18	25	0,010
5	18	20	0,036	10	18	25	0,020	15	20	25	0,009	20	20	20	0,0002
5	20	16	0,001	10	20	20	0,001	15	22	25	0,049	20	20	25	0,006
5	20	20	0,026	10	20	25	0,014	20	2	2	0,002	20	22	25	0,003
5	24	20	0,004	10	22	25	0,070								
$\alpha = 0{,}95$															
5	2	2	0,004	5	12	20	0,050	10	8	8	0,001	15	4	12	0,040
5	2	3	0,030	5	12	25	0,070	10	8	10	0,010	15	4	16	0,050
5	2	4	0,052	5	14	16	0,010	10	8	12	0,020	15	4	20	0,061
5	2	5	0,080	5	14	20	0,033	10	8	16	0,036	15	4	25	0,075
5	2	6	0,100	5	14	25	0,058	10	8	20	0,050	15	6	6	0,0006
5	2	8	0,130	5	16	16	0,001	10	8	25	0,065	15	6	8	0,0085
5	2	10	0,180	5	16	20	0,020	10	10	10	0,001	15	6	10	0,019
5	2	12	0,190	5	16	25	0,044	10	10	12	0,0075	15	6	12	0,024
5	2	16	0,210	5	18	20	0,008	10	10	16	0,022	15	6	16	0,036
5	2	20	0,250	5	18	25	0,031	10	10	20	0,036	15	6	20	0,047
5	2	25	0,270	5	20	20	0,001	10	10	25	0,050	15	6	25	0,059
5	4	4	0,002	5	20	25	0,020	10	12	12	0,001	15	8	8	0,0005

Окончание таблицы 213

m	n	k	b	m	n	k	b	m	n	k	b	m	n	k	b
$\alpha = 0{,}95$															
5	4	5	0,020	5	22	25	0,012	10	12	16	0,012	15	8	10	0,0065
5	4	6	0,033	10	2	2	0,002	10	12	20	0,026	15	8	12	0,014
5	4	8	0,060	10	2	3	0,010	10	12	25	0,040	15	8	16	0,025
5	4	10	0,090	10	2	4	0,029	10	14	16	0,0045	15	8	20	0,035
5	4	12	0,118	10	2	5	0,040	10	14	20	0,018	15	8	25	0,045
5	4	16	0,150	10	2	6	0,050	10	14	25	0,030	15	10	10	0,0004
5	4	20	0,180	10	2	8	0,070	10	16	20	0,010	15	10	12	0,006
5	4	25	0,200	10	2	10	0,080	10	16	25	0,022	15	10	16	0,018
5	6	6	0,002	10	2	12	0,100	10	18	20	0,004	15	10	20	0,026
5	6	8	0,022	10	2	16	0,130	10	18	25	0,017	15	10	25	0,036
5	6	10	0,045	10	2	20	0,150	10	20	25	0,010	15	12	12	0,0003
5	6	12	0,061	10	2	25	0,160	10	22	25	0,005	15	12	16	0,009
5	6	16	0,097	10	4	4	0,002	15	2	2	0,0018	15	12	20	0,018
5	6	20	0,130	10	4	5	0,010	15	2	3	0,010	15	12	25	0,028
5	6	25	0,160	10	4	6	0,016	15	2	4	0,020	15	14	16	0,003
5	8	8	0,001	10	4	8	0,030	15	2	5	0,030	15	14	20	0,013
5	8	10	0,016	10	4	10	0,046	15	2	6	0,040	15	14	25	0,022
5	8	12	0,033	10	4	12	0,058	15	2	8	0,050	15	16	16	0,0002
5	8	16	0,065	10	4	16	0,078	15	2	10	0,060	15	16	20	0,0075
5	8	20	0,090	10	4	20	0,095	15	2	12	0,070	15	16	25	0,013
5	8	25	0,130	10	4	25	0,110	15	2	16	0,080	15	18	20	0,003
5	10	10	0,001	10	6	6	0,001	15	2	20	0,090	15	18	25	0,013
5	10	12	0,015	10	6	8	0,012	15	2	25	0,100	15	20	20	0,0001
5	10	16	0,044	10	6	10	0,024	15	4	4	0,0008	15	20	25	0,008
5	10	20	0,070	10	6	12	0,036	15	4	5	0,006	15	22	25	0,004
5	10	25	0,090	10	6	16	0,051	15	4	6	0,010				
5	12	12	0,001	10	6	20	0,070	15	4	8	0,025				
5	12	16	0,025	10	6	25	0,080	15	4	10	0,030				

4.4.3.2. Прогнозные интервалы в задачах испытаний на надежность

В [549] рассмотрена задача прогнозирования r-й порядковой статистики x_r в выборке объема n, основанная на величине первых k порядковых наблюдений из выборки ($k < r \leqslant n$). Доверительный интервал для x_r может быть использован для прогнозирования оставшегося времени испытаний на долговечность. Показано [549], что статистика $U = \dfrac{x_r - x_k}{\sum\limits_{i=1}^{k} x_i + (-k)x_k}$ имеет F-распределение с $f_1 = 2$ и $f_2 = 2(r-1)$ степенями свободы. Следовательно, двусторонний интервал для x_r с доверительной вероятностью α будет

$$x_k + F_{\frac{1-\alpha}{2}} \left[\sum_{i=1}^{k} x_i + (n-k)\,x_k \right] \leqslant x_r \leqslant x_k + F_{\frac{1+\alpha}{2}} \left[\sum_{i=1}^{k} x_i + (n-k)\,x_k \right].$$

Для односторонних интервалов вместо $\dfrac{1+\alpha}{2}$ или $\dfrac{1-\alpha}{2}$ используются величины α или $(1-\alpha)$ соответственно.

Хан [550] рассмотрел задачу прогнозирования будущих выборок из экспоненциального распределения. Задача формулируется следующим образом. По имеющимся данным моментов отказов m изделий необходимо найти нижний доверительный

интервал с вероятностью α для среднего времени до отказа для k будущих выборок объема n. Если $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$, то нижняя доверительная граница для средних k выборок объема n будет равна $\bar{x}c(k,m,n,\alpha)$, где $c(k,m,n,\alpha)$ — коэффициенты оценки. Справедливы аппроксимации

$$c(k,m,n,\alpha) \approx \frac{1}{F\left(2m; 2n; (1-\alpha)^{\frac{1}{k}}\right)}; \quad c(k,m,n,\alpha) \approx \frac{1}{F\left(2m; 2n; 1-\frac{\alpha}{k}\right)},$$

где $F(a,b,\gamma)$ — γ-квантиль распределения Фишера (см. раздел 1.1.10) с $f_1 = a$ и $f_2 = b$ степенями свободы.

Первая аппроксимация точнее, однако весьма затруднен поиск $(1-\alpha)^{\frac{1}{k}}$-квантилей F-распределения (имеющиеся таблицы недостаточны). Вторая аппроксимация менее точна, но может быть получена с помощью имеющихся таблиц F-распределения. Некоторые значения коэффициентов $c(k,n,m,\alpha)$, полученные с помощью первой аппроксимации, приведены в табл. 214.

Таблица 214

Значения $c(k,n,m,\alpha)$ для прогнозирования при $\alpha = 0{,}95$ нижней границы средней наработки в k выборках объема n по m ранним моментам отказов [550]

m	$n=2$					$n=5$					$n=10$				
	k					k					k				
	2	3	5	8	12	2	3	5	8	12	2	3	5	8	12
3	0,8	1,1	1,4	1,5	1,7	1,9	2,8	3,6	4,0	4,7	3,3	4,9	6,5	7,9	9,0
5	0,5	0,7	0,9	0,8	0,9	1,1	1,6	2,1	2,5	2,7	2,0	3,0	4,1	4,9	5,6
10	0,2	0,4	0,4	0,4	0,5	0,5	0,8	1,0	1,1	1,3	1,0	1,5	2,0	2,3	2,6
20	0,1	0,2	0,2	0,2	0,2	0,3	0,4	0,5	0,5	0,6	0,5	0,7	0,9	1,0	1,2

В заключение приведем ряд имеющих практический интерес соотношений, основанных на гипергеометрическом распределении (см. раздел 1.2.6). Вероятность того, что среди будущих m наблюдений не менее c величин будут превышать r-й член упорядоченного по возрастанию ряда в выборке объема n, равна

$$p = \sum_{i=\max(0,c+r-1-m)}^{r-1} \frac{(c+r-1)!n!}{(c+r-1-i)!(n-i)!i!}\,\frac{(m+n-c-r+1)!m!}{(m+n)!(m-c-r+1+i)!}.$$

Вероятность p^* того, что среди будущих m наблюдений не менее c величин будут заключены между крайними значениями ряда, построенного по выборке объема n, равна

$$p^* = 1 - \frac{(m-c+2)!n!}{(m-c+1)!\,(n-1)!}\,\frac{(n+c-2)!m!}{(m+n)!\,(c-1)!}.$$

Задача 315. *В результате испытаний $n = 10$ приборов первые $k = 7$ отказов получены в моменты времени t_i: 10, 12, 18, 24, 31, 35, 41. Необходимо найти доверительные интервалы, в которых находится момент отказа девятого прибора ($r = 9$), при доверительной вероятности $\alpha = 0{,}95$.*

Имеем $t_k = t_7 = 41$. Из таблиц F-распределения (либо используя аппроксимации из раздела 1.1.10) имеем (при $f_1 = 2$ и $f_2 = 2\cdot(9-1) = 16$)

$$F_\alpha = F_{0,95} = 3{,}63; \quad F_{1-\alpha} = F_{0,05} = 0{,}0515; \quad F_{\frac{1+\alpha}{2}} = F_{0,975} = 4{,}96;$$

$$F_{\frac{1}{1-\alpha}} = F_{0,025} = 0{,}0253.$$

Далее находим $\sum_{i=1}^{7} t_i + (10-7)\,41 = 171 + 13 = 294$. Для двустороннего интервала имеем

$$41 + 294 \cdot 0{,}253 = 48{,}438 \leqslant t_9 \leqslant 1419{,}86 = 41 + 294 \cdot 4{,}69.$$

Для односторонних интервалов имеем

$$t_9 \geqslant 41 + 294 \cdot 0{,}0515 = 56{,}141; \quad t_9 \leqslant 41 + 294 \cdot 3{,}63 = 1108{,}22.$$

Задача 316. *Для $m = 10$ изделий получены моменты отказов t_i: 158, 171, 192, 211, 241, 256, 278, 292, 312, 341. Необходимо с достоверностью $\alpha = 0{,}95$ найти нижний доверительный интервал для средней наработки на отказ в $k = 5$ выборках объема $n = 10$ каждая.*

Имеем для $\alpha = 0{,}95$ из табл. 214 коэффициент $c(5, 10, 10) = 2{,}0$.

Далее $\bar{t} = \frac{1}{10} \cdot \sum_{i=1}^{10} t_i = 245{,}2$, и нижняя граница оценки равна

$$\bar{t} \cdot c(k, n, m, \alpha) = 245,2 \cdot 2,0 = 490,4.$$

Следовательно, с вероятностью 0,95 средняя наработка на отказ в 5 будущих выборках не будет превышать 490,4.

Задача 317. *Для данных задачи 316 определить вероятность того, что среди будущих трех наблюдений не менее двух будут превышать по величине 5-е наблюдение $t_5 = 241$ в выборке.*

Имеем $n = 10$, $r = 5$, $m = 3$ и $c = 2$. Далее вычисляем

$$p = \sum_{i=\max(0,c+r-1-m)}^{r-1} \frac{(c+r-1)!n!}{(c+r-1-i)!\,(n-i)!i!} \cdot \frac{(m+n-c-r+1)!m!}{(m+n)!\,(m-c-r+1+i)!} =$$

$$= \sum_{i=3}^{4} \frac{6!10!}{(6-i)!\,(10-i)!i!} \cdot \frac{7!3!}{13!\,(i-3)!} = 0{,}437.$$

Следовательно, с вероятностью $p = 0{,}437$ среди трех будущих наблюдений не менее двух будут превышать величину $t_5 = 241$.

Задача 318. *Для данных задачи 316 вычислить вероятность того, что среди будущих $m = 3$ наблюдений не менее $c = 2$ величин будут заключены между крайними значениями выборки объема $n = 10$ (т. е. между $t_1 = 158$ и $t_{10} = 341$).*

Имеем

$$p^* = 1 - \frac{(3-2+2)!10!}{(3-2+1)!9!} \cdot \frac{(10+2-2)!3!}{(10+3)!\,(2-i)!} = 0{,}895.$$

Следовательно, с вероятностью $p^* = 0{,}895$ из будущих трех наблюдений не менее двух будут находиться между минимальным $t_1 = 158$ и максимальным $t_{10} = 341$ значениями выборки объема $n = 10$.

ГЛАВА 5

МЕТОДЫ ИССЛЕДОВАНИЯ СВЯЗЕЙ МЕЖДУ СЛУЧАЙНЫМИ ВЕЛИЧИНАМИ

Общие положения. В предыдущих главах были рассмотрены различные методы и приемы математической статистики, позволяющие оценить параметры статистических совокупностей, сравнивать их между собой. При этом, как правило, предполагалась взаимная независимость сравниваемых совокупностей. В настоящей главе рассматриваются вопросы оценки связей между статистическими совокупностями.

Подробно описываются основы и методы дисперсионного, корреляционного и регрессионного анализов, являющиеся последовательными ступенями при исследовании связей между случайными величинами.

Методами дисперсионного анализа устанавливается наличие влияния заданного фактора на изучаемый процесс, отображаемый наблюдаемой статистической совокупностью выборочных данных. Корреляционный анализ позволяет оценить силу такой связи, а методами регрессионного анализа можно выбрать конкретную математическую модель и оценить адекватность отражения ею установленной взаимосвязи случайных величин.

В последние годы стремительно развивается самостоятельное прикладное направление математической статистики — математическая теория активного эксперимента. Базируясь на комбинации методов дисперсионного и регрессионного анализов, методы математического планирования эксперимента дополняют их.

В настоящей главе даются основные понятия и определения математической теории планирования активного эксперимента применительно к изучению механизма наблюдаемого процесса и построению его статистической модели.

Автор счел уместным изложить в настоящей главе основные сведения по технике контрольных карт, как самой распространенной форме применения математической статистики в производстве при предупредительном статистическом контроле качества продукции. Приводится краткое описание известных методов и примеры их применения.

5.1. Дисперсионный анализ

Дисперсионный анализ является статистическим методом анализа результатов наблюдений, зависящих от различных одновременно действующих факторов, с целью выбора наиболее значимых факторов и оценки их влияния на исследуемый процесс.

Влияние различных факторов на изучаемые случайные величины (например, влияние технологического способа изготовления или режима нагрузки на долговечность технического изделия) приводит к изменению значений параметров распределения вероятностей этих величин — среднего, дисперсии или моментов более высокого порядка.

С помощью дисперсионного анализа устанавливаются изменения дисперсии результатов эксперимента при изменении уровней изучаемого фактора. Если дисперсии будут отличаться значимо, то следует вывод о значимом влиянии фактора на среднее значение наблюдаемой случайной величины.

Классические методы дисперсионного анализа основываются на следующих предпосылках: распределение исходных случайных величин нормально; дисперсии экспериментальных данных одинаковы для всех условий эксперимента (т. е. для экспериментов, выполненных на различных уровнях изучаемого фактора).

Поэтому при проведении дисперсионного анализа следует предварительно проверить нормальность распределения изучаемой случайной величины (методами, изложенными в разделе 2.2) и неразличимость дисперсий изучаемых совокупностей (методами, изложенными в разделе 4.1.1.4).

Подробно теория дисперсионного анализа изложена в [20, 401].

5.1.1. Классический дисперсионный анализ нормально распределенных случайных величин

5.1.1.1. Однофакторный дисперсионный анализ

Предположим, что анализируется влияние фактора A, изучаемого на k уровнях $(A_1, A_2, \ldots, A_k)$. На каждом уровне A_i проведены n наблюдений $(x_{i1}, x_{i2}, \ldots, x_{in})$. Следовательно, на всех k уровнях фактора A произведены kn наблюдений.

Поясним суть и последовательность проведения дисперсионного анализа. Расположим экспериментальные данные в виде таблицы:

Номер наблюдения	Уровни фактора A					
	A_1	A_2	...	A_i	...	A_k
1	x_{11}	x_{21}	...	x_{i1}	...	x_{k1}
2	x_{12}	x_{22}	...	x_{i2}	...	x_{k2}
.	.	.	...	.	...	.
.	.	.	...	.	...	.
.	.	.	...	.	...	.
j	x_{1j}	x_{2j}	...	x_{ij}	...	x_{kj}
.	.	.	...	.	...	.
.	.	.	...	.	...	.
.	.	.	...	.	...	.
n	x_{1n}	x_{2n}	...	x_{in}	...	x_{kn}
$\sum$	X_1	X_2	...	X_i	...	X_k

Рассмотрим оценки различных дисперсий, возникающие при анализе таблицы результатов наблюдений. Для дисперсии, характеризующей изменение данных на уровне A_i (по строкам таблицы), имеем

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2 = \frac{1}{n-1}\left[\sum_{j=1}^{n}x_{ij}^2 - \frac{1}{n}\left(\sum_{j=1}^{n}x_{ij}\right)^2\right].$$

Из предпосылок дисперсионного анализа следует, что должно иметь место равенство $S_1^2 = S_2^2 = \ldots = S_k^2$, что проверяется соответствующим критерием сравнения (см. раздел 4.1.1.4).

При выполнении условия $S_i^2 = \text{const}$ при $i = 1, 2, \ldots, k$ (что, повторяем весьма настойчиво, обязательно), находим оценку дисперсии, характеризующей рассеяние

значений x_{ij} вне влияния фактора A, по формуле

$$S_0^2 = \frac{1}{k}\sum_{i=1}^{k} S_i^2 \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2 = \frac{1}{k(n-1)}\left[\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{n}\sum_{i=1}^{k}\left(\sum_{j=1}^{n} x_{ij}\right)^2\right].$$

Легко видеть, что, если при оценке S_i^2 мы имеем $(n-1)$ степеней свободы, то оценка S_0^2 имеет $k\,(n-1)$ степеней свободы.

Общая выборочная дисперсия всех наблюдений равна

$$S^2 = \frac{1}{kn-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{\bar{x}})^2, \quad \text{где} \quad \bar{\bar{x}} = \frac{1}{k}\sum_{i=1}^{k}\bar{x}_i; \quad \bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}.$$

Следовательно,

$$S^2 = \frac{1}{kn-1}\left[\sum_{i=1}^{k}\sum_{j=1}^{k} x_{ij}^2 - \frac{1}{kn}\left(\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}\right)^2\right].$$

Введем теперь оценку дисперсии S_A^2, характеризующей изменение средних $\bar{x}_i$, связанное с влиянием фактора A:

$$S_A^2 = \frac{n}{k-1}\sum_{i=1}^{k}(\bar{x}_i-\bar{\bar{x}})^2.$$

Очевидно, что при оценке S_A^2 используется $(k-1)$ степеней свободы. Теперь проверка влияния фактора A на изменение средних может быть сведена к сравнению дисперсий S_A^2 и S_0^2. Влияние фактора A признается значимым, если значимо отношение $\frac{S_A^2}{S_0^2}$. Отношение $\frac{S_A^2}{S_0^2}$. признается значимым с вероятностью α, если

$$\frac{S_A^2}{S_0^2} > F_\alpha[k-1;\, k\,(n-1)],$$

где $F_\alpha(f_1, f_2)$ — α-квантиль F-распределения с f_1 и f_2 степенями свободы (для нахождения $F_\alpha(f_1, f_2)$ могут быть использованы либо специальные таблицы, например, из [24, 25, 29], либо аппроксимации из раздела 1.1.10).

Если влияние фактора A признается незначимым (т. е., когда $\frac{S_A^2}{S_0^2} \leqslant F_\alpha(f_1, f)$), то для оценки дисперсии S_0^2 может быть использована более точная оценка S^2, имеющая $(kn-1)$ степеней свободы против $k\,(n-1)$ для S_0^2.

Для упрощения вычислений приведем алгоритм их выполнения. Вычисляем последовательно суммы

$$Q_1 = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2; \quad Q_2 = \frac{1}{n}\sum_{i=1}^{k} X_i^2; \quad Q_3 = \frac{1}{kn}\left(\sum_{i=1}^{k} X_i\right)^2.$$

Далее находим

$$S_0^2 = \frac{Q_1 - Q_2}{k\,(n-1)}; \quad S_A^2 = \frac{Q_2 - Q_3}{k-1}.$$

Сравнением S_A^2 и S_0^2 устанавливаем наличие влияния фактора A.

Если $\dfrac{k\,(n-1)}{k-1}\dfrac{Q_2-Q_3}{Q_1-Q_2} > F_\alpha[k-1; k\,(n-1)]$, то влияние фактора A признается значимым. В ином случае всю выборку наблюдений можно считать однородной с общей дисперсией $S^2 = \dfrac{Q_1-Q_3}{kn-1}$.

Ранее (см. раздел 4.1.1.2), однофакторный дисперсионный анализ использовался для проверки гипотезы о равенстве нескольких выборочных средних из нормального распределения. Однофакторный дисперсионный анализ и сравнение нескольких средних являются различными приемами решения одной и той же задачи. Когда на различных уровнях фактора A проводятся различные количества наблюдений, формулы дисперсионного анализа имеют вид (n_i — число экспериментов на уровне A_i, $N = \sum\limits_{i=1}^{k} n_i$):

$$Q_1 = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2; \quad Q_2 = \sum_{i=1}^{k} \frac{X_i^2}{n_i}; \quad Q_3 = \frac{1}{N}\left(\sum_{i=1}^{k} X_i\right)^2;$$

$$S_0^2 = \frac{Q_1-Q_2}{N-k}; \quad S_A^2 = \frac{Q_2-Q_3}{k-1}.$$

Отношение $\dfrac{S_A^2}{S_0^2}$ сравнивается с $F_\alpha(k-1; N-k)$.

Задача 319. *Провести дисперсионный анализ данных, представленных таблицей, при доверительной вероятности* $\alpha = 0{,}95$:

i	Уровни фактора A_i				
	A_1	A_2	A_3	A_4	A_5
1	3,2	2,6	2,9	3,6	3,0
2	3,1	3,1	2,6	3,4	3,4
3	3,1	2,7	3,0	3,2	3,2
4	2,8	2,9	3,1	3,3	3,5
5	3,3	2,7	3,0	3,5	2,9
6	3,0	2,8	2,8	3,3	3,1
$\sum$	18,5	16,8	17,4	20,3	19,1

Вычисляем

$$Q_1 = \sum_{i=1}^{5}\sum_{j=1}^{6} x_{ij}^2 = 284{,}8; \quad Q_2 = \frac{1}{6}\cdot\sum_{i=1}^{5} X_i^2 = \frac{1}{6}\cdot\left(18{,}5 + 16{,}8^2 + \ldots + 19{,}1^2\right) = 284{,}025;$$

$$Q_3 = \frac{1}{5\cdot 6}\cdot\left(\sum_{i=1}^{5} X_i\right)^2 = \frac{1}{30}\cdot(18{,}5+16{,}8+17{,}4+20{,}3+19{,}1)^2 = 282{,}747.$$

Далее вычисляем дисперсии

$$S_0^2 = \frac{284{,}87-284{,}025}{5\cdot(6-1)} = 0{,}0338; \quad S_A^2 = \frac{284{,}025-282{,}747}{5-1} = 0{,}319; \quad \frac{S_A^2}{S_0^2} = \frac{0{,}319}{0{,}0338} = 9{,}45.$$

Из таблиц для $f_1 = k-1 = 4$ и $f_2 = k\cdot(n-1) = 25$ находим $F_{0,95}(4;25) = 2{,}8$. Так как $\dfrac{S_A^2}{S_0^2} = 9{,}45 > F_{0,95}(4;25) = 2{,}8$, влияние фактора A на поведение наблюдаемых случайных величин следует признать значимым.

5.1.1.2. Двухфакторный дисперсионный анализ

Рассмотренный ранее однофакторный дисперсионный анализ обладает информативностью, не большей, чем методы множественного сравнения средних (см. раздел 4.1.1.2). Информативность дисперсионного анализа возрастает при одновременном изучении влияния нескольких факторов.

Рассмотрим случай, когда анализируется влияние одновременно двух факторов A и B на уровнях $A_1, A_2, \ldots, A_k$ и $B_1, B_2, \ldots, B_m$ соответственно. Пусть результаты эксперимента представлены таблицей:

B	A						$\sum$
	A_1	A_2	$\ldots$	A_i	$\ldots$	A_k	
B_1	x_{11}	x_{21}	$\ldots$	x_{i1}	$\ldots$	x_{k1}	$X_{1'}$
B_2	x_{12}	x_{22}	$\ldots$	x_{i2}	$\ldots$	x_{k2}	$X_{2'}$
.	.	.	$\ldots$	.	$\ldots$	$\ldots$	.
.	.	.	$\ldots$	.	$\ldots$	$\ldots$	.
.	.	.	$\ldots$	.	$\ldots$	$\ldots$	.
B_j	x_{ij}	x_{2j}	$\ldots$	x_{ij}	x_{kj}	x_{kj}	$X_{j'}$
.	.	.	$\ldots$	.	$\ldots$	$\ldots$	.
.	.	.	$\ldots$	.	$\ldots$	$\ldots$	.
.	.	.	$\ldots$	.	$\ldots$	$\ldots$	.
B_m	x_{1m}	x_{2m}	$\ldots$	x_{im}	$\ldots$	x_{km}	$X_{m'}$
$\sum$	X_1	X_2	$\ldots$	X_i	$\ldots$	X_k	

Дисперсионный анализ для двухфакторных таблиц проводится в следующей последовательности. Вычисляются суммы

$$Q_1 = \sum_{i=1}^{k}\sum_{j=1}^{m} x_{ij}^2; \quad Q_2 = \frac{1}{m}\sum_{i=1}^{k} X_i^2; \quad Q_3 = \frac{1}{k}\sum_{j=1}^{m} X_{j'}^2;$$

$$Q_4 = \frac{1}{mk}\left(\sum_{i=1}^{k} X_i\right)^2 = \frac{1}{mk}\left(\sum_{j=1}^{m} X_{j'}\right)^2.$$

Далее находятся оценки дисперсий

$$S_0^2 = \frac{Q_1 + Q_4 - Q_2 - Q_3}{(k-1)(m-1)}; \quad S_A^2 = \frac{Q_2 - Q_4}{k-1}; \quad S_B^2 = \frac{Q_3 - Q_4}{m-1}.$$

Если $\dfrac{S_A^2}{S_0^2} > F_\alpha(f_1, f_2)$, где $f_1 = k-1$ и $f_2 = (k-1)(m-1)$, то влияние фактора A с достоверностью α признается значимым.

Аналогично значимым признается влияние фактора B, если

$$\frac{S_B^2}{S_0^2 > F\alpha}(f_1, f_2), \quad \text{где} \quad f_1 = m-1 \quad \text{и} \quad f_2 = (k-1)(m-1).$$

Приведенный анализ предполагает независимость факторов A и B. Если они зависимы, то взаимодействие факторов $C = AB$ также является фактором, которому соответствует своя дисперсия. Для того, чтобы выделить такое взаимодействие, необходимы параллельные наблюдения в каждой клетке таблицы, т. е. при каждом сочетании факторов A и B на уровнях A_i и B_j соответственно необходимо не одно наблюдение, а серия наблюдений $x_{ij1}, x_{ij2}, \ldots, x_{ijn}$. Пусть x_{ij} теперь является

средним из n наблюдений, т. е. $x_{ij} = \dfrac{1}{n}\sum\limits_{\nu=1}^{n} x_{ij\nu}$. Для оценки влияния взаимодействия факторов AB вычисляем дополнительную сумму

$$Q_5 = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{\nu=1}^{n} x_{ij\nu}^2.$$

Далее анализ проводится, как и ранее, с той лишь разницей, что в клетках таблицы вместо отдельных значений $x_{ij\nu}$ используются их средние значения x_{ij}.

Вычисляется дисперсия

$$S_{AB}^2 = \frac{Q_5 - nQ_1}{mk\,(n-1)},$$

и проверяется значимость взаимодействия факторов AB критерием

$$\frac{nS_0^2}{S_{AB}^2} > F_\alpha(f_1, f_2), \quad \text{где} \quad f_1 = (k-1)(m-1) \quad \text{и} \quad f_2 = mk(n-1).$$

С добавлением каждого нового фактора принципиальная основа дисперсионного анализа не изменяется, но существенно усложняются формулы и таблицы для расчетов. Подробное изложение прикладных методов дисперсионного анализа для случая трех, четырех и более факторов с анализом различных практических ситуаций содержится в [20, 559].

Задача 320. *Провести двухфакторный дисперсионный анализ данных, представленных следующей таблицей, при доверительной вероятности* $\alpha = 0{,}95$:

B	A								
	A_1			A_2			A_3		
B_1	3,6	3,8	4,1	2,9	3,1	3,0	2,6	2,5	2,9
B_2	4,2	4,0	4,1	3,3	2,9	3,2	3,7	3,5	3,6
B_3	3,8	3,5	3,6	3,6	3,7	3,5	3,2	3,0	3,4
B_4	3,4	3,2	3,2	3,4	3,6	3,5	3,6	3,8	3,7

Заменяя в клетках таблицы серии значений их средними, получаем следующую таблицу:

B	A			$\sum$
	A_1	A_2	A_3	
B_1	3,83	3,00	2,67	9,50
B_2	4,10	3,13	3,60	10,83
B_3	3,63	3,60	3,20	10,43
B_4	3,27	3,50	3,70	10,47
$\sum$	14,83	13,23	13,17	41,23

Используя данные таблицы, вычисляем суммы

$$Q_1 = \sum_{i=1}^{3}\sum_{j=1}^{4} x_{ij}^2 = 143{,}34; \quad Q_2 = \frac{1}{4}\cdot\sum_{i=1}^{3} X_i^2 = 142{,}102675; \quad Q_3 = \frac{1}{3}\cdot\sum_{j=1}^{4} X_j^2 = 141{,}98157;$$

$$Q_4 = \frac{1}{4\cdot 3}\cdot\left(\sum_{i=1}^{3} X_i\right)^2 = 141{,}6594; \quad Q_5 = \sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{\nu=1}^{3} x_{ij\nu}^2 = 430{,}79.$$

Далее вычисляем:

$$S_0^2 = \frac{Q_1 + Q_4 - Q_2 - Q_3}{(k-1)\cdot(m-1)} = \frac{143{,}3745 + 141{,}6594 - 142{,}102675 - 141{,}98157}{2\cdot 3} = 0{,}1582;$$

$$S_A^2 = \frac{Q_2 - Q_4}{k-1} = \frac{142{,}3745 - 141{,}6594}{2} = 0{,}223675;$$

$$S_B^2 = \frac{Q_3 - Q_4}{m-1} = \frac{141{,}98157 - 141{,}6594}{3} = 0{,}10739;$$

$$S_{AB}^2 = \frac{Q_5 - n\cdot Q_1}{mk\cdot(n-1)} = \frac{430{,}79 - 3\cdot 143{,}3745}{4\cdot 3\cdot 2} = 0{,}02777;$$

$$\frac{S_A^2}{S_0^2} = \frac{0{,}223675}{0{,}1582} = 1{,}41; \quad \frac{S_B^2}{S_0^2} = \frac{0{,}10739}{0{,}1582} = 0{,}679; \quad \frac{n\cdot S_0^2}{S_{AB}^2} = \frac{3\cdot 0{,}1582}{0{,}02777} = 17{,}09.$$

Из таблиц (см. раздел 1.1.10) имеем

$$F_{0,95}[k-1;(k-1)\cdot(m-1)] = F_{0,95}(2;6) = 5{,}1; \quad F_{0,95}[m-1;(k-1)\cdot(m-1)] = 4{,}8;$$
$$F_{0,95}[(k-1)\cdot(m-1)\,;mk\cdot(n-1)] = F_{0,95}(6;24) = 2{,}5.$$

Сравнивая, получаем

$$\frac{S_A^2}{S_0^2} = 1{,}41 < F_{0,95}(2;6) = 5{,}1; \quad \frac{S_B^2}{S_0^2} = 0{,}679 < F_{0,95}(3;6) = 4{,}8;$$

$$\frac{n\cdot S_0^2}{S_{AB}^2} = 17{,}09 > F_{0,95}(6;24) = 2{,}5.$$

Следовательно, влияние факторов A и B должно быть признано незначимым. Однако существенно значимым является взаимодействие факторов A и B. Это поучительный пример эффективности двухфакторного дисперсионного анализа по сравнению с простым последовательным повторением однофакторных экспериментов.

5.1.2. Дисперсионный анализ с использованием размахов

Рассмотрим схему двухфакторного дисперсионного анализа с k уровнями фактора A и m уровнями фактора B (при каждом сочетании уровней факторов одно наблюдение или среднее из нескольких наблюдений). В [119] рассмотрено применение „стьюдентизированного" размаха, как статистики для выявления значимости влияния изучаемых факторов

$$q = \frac{c\sqrt{m}\left(\max\limits_{1\leqslant i\leqslant k}\bar{x}_i - \min\limits_{1\leqslant i\leqslant k}\bar{x}_i\right)}{\bar{\omega}}.$$

Статистика q распределена как статистика „стьюдентизированного" размаха (см. раздел 4.1.1.2.2), критические точки которой приведены в табл. 123 для различных m (в табл. 123 следует вместо n использовать m) и f. Количества эквивалентных степеней свободы f для нашего случая и значения коэффициента c приведены в табл. 215.

Средний размах $\bar{\omega}$ вычисляется следующим образом. Для каждого уровня A_i находим

$$\bar{x}_i = \frac{1}{m}\sum_{i=1}^{m} x_{ij}, \quad i = 1, 2, \ldots, k.$$

Затем для всех уровней фактора B при i-м уровне фактора A вычисляем m разностей $\Delta x_{ij} = x_{ij} - \bar{x}_i$, $j = 1, 2, \ldots, m$. Далее находим

$$\omega_j = \max_{1\leqslant i\leqslant k}\Delta x_{ij} - \min_{1\leqslant i\geqslant k}\Delta x_{ij} \quad \text{и} \quad \bar{\omega} = \frac{1}{m}\sum_{j=1}^{m}\omega_i.$$

Таблица 215

Масштабный коэффициент c и эффективное число степеней свободы f для двухфакторного дисперсионного анализа [119]

m	$k=2$		$k=3$		$k=4$		$k=5$	
	f	c	f	c	f	c	f	c
2	1,00	1,00	2,00	1,35	2,90	1,58	3,80	1,75
3	1,90	1,05	3,70	1,48	5,60	1,76	7,40	1,96
4	2,70	1,07	5,40	1,54	8,20	1,84	11,00	2,06
5	3,60	1,08	7,20	1,57	10,90	1,88	14,60	2,12
6	4,50	1,09	8,90	1,59	13,60	1,91	18,20	2,15
7	5,40	1,09	10,70	1,61	16,30	1,93	21,80	2,18
8	6,30	1,10	12,50	1,62	19,00	1,95	25,40	2,20
9	7,10	1,10	14,30	1,63	21,70	1,96	29,00	2,21
10	8,10	1,10	16,10	1,63	24,40	1,97	32,60	2,22
20	16,70	11,10	33,90	1,66	51,50	2,02	68,80	2,28

m	$k=6$		$k=7$		$k=8$		$k=9$	
	f	c	f	c	f	c	f	c
2	4,70	1,89	5,50	2,00	6,30	2,10	7,00	2,18
3	9,30	2,12	11,30	2,26	13,40	2,37	15,70	2,46
4	13,90	2,23	16,90	2,38	20,10	2,50	23,60	2,60
5	18,50	2,30	22,40	2,45	26,60	2,57	31,10	2,68
6	23,00	2,34	27,90	2,49	33,00	2,62	38,30	2,73
7	27,60	2,37	33,30	2,52	39,30	2,65	45,40	2,76
8	32,10	2,39	38,70	2,55	45,60	2,68	52,50	2,79
9	36,60	2,41	44,00	2,57	51,80	2,70	59,60	2,81
10	41,00	2,42	49,30	2,58	57,90	2,71	66,60	2,83
20	86,00	2,48	103,00	2,64	119,00	2,78	134,00	2,90

Превышение выборочной статистикой q ее критического значения $q_\alpha(m, f)$ (из табл. 123) приводит к признанию значимости влияния изучаемых факторов.

Задача 321. *Провести дисперсионный анализ в условиях задачи* 320 *с помощью размахов.*

Таблица данных имеет следующий вид:

B	A_1	A_2	A_3
B_1	3,83	3,00	2,70
B_2	4,10	3,13	3,50
B_3	3,63	3,60	3,20
B_4	3,27	3,50	3,70

Имеем $k = 3$, $m = 4$. Из табл. 215 для $k = 3$, $m = 4$ находим $c = 1{,}54$ и $f = 5{,}4$.

Далее средние по столбцам (при уровне a_i фактора A) равны

$$\bar{x}_1 = \frac{3{,}83 + 4{,}10 + 3{,}63 + 3{,}27}{4} = 3{,}7075; \quad \bar{x}_2 = \frac{3{,}00 + 3{,}13 + 3{,}60 + 3{,}50}{4} = 3{,}3075;$$

$$\bar{x}_3 = \frac{2{,}70 + 3{,}50 + 3{,}20 + 3{,}70}{4} = 3{,}275.$$

Следовательно, $\max\limits_{1\leqslant i\leqslant 3} \bar{x}_i - \min\limits_{1\leqslant i\leqslant 3} \bar{x}_i = \bar{x}_1 - \bar{x}_3 = 3{,}7075 - 3{,}275 = 0{,}4325.$

Теперь вычислим $m = 4$ разностей $\Delta x_{ij} = x_{ij} - \bar{x}_i$ для всех $j = 1, 2, \ldots, m$. Таблица этих разностей для нашего случая имеет вид.

B	A			ω_j
	A_1	A_2	A_3	
B_1	0,1225	−0,3075	−0,5750	0,6975
B_2	0,3925	−0,1775	0,2250	0,5700
B_3	−0,0775	0,2925	−0,0750	0,3700
B_4	−0,4375	0,1925	0,4250	0,8625

Для каждого j значения $\omega_j = \max\limits_{1 \leqslant i \leqslant 3} \Delta x_{ij} - \min\limits_{1 \leqslant i \leqslant 3} \Delta x_{ij}$ приведены в последнем столбце таблицы. Далее находим

$$\bar{\omega} = \frac{0{,}6975 + 0{,}570 + 0{,}37 + 0{,}8625}{4} = 0{,}625; \quad q = 1{,}54 \cdot \sqrt{4} \cdot \frac{0{,}4325}{0{,}625} = 2{,}13.$$

Из табл. 123 находим для $m = 4$ (входим в таблицу с $n = 3$) $f = 5{,}4$ (используем интерполяцию) и при $\alpha = 0{,}95$ находим $q_{0,95}(4; 5{,}4) = 5{,}2$.

Так как $q = 2{,}13 < q_{0,95}(4; 5{,}4) = 5{,}2$, этот критерий не выявил влияния факторов A и B на наблюдаемый процесс. К сожалению, выделить влияние взаимодействия факторов A и B с помощью этого критерия невозможно.

5.1.3. Непараметрический дисперсионный анализ

Использование для однофакторного дисперсионного анализа вместо значений случайных величин их рангов, назначенных определенным образом, рассмотрено в разделе 4.2.1.2.15, где приведена подробная информация о критерии Фридмана–Кендалла–Бэбингтона Смита, применяемом в этом случае. Мы отсылаем читателя к разделу 4.2.1.2.15, где этот критерий рассмотрен применительно к задаче сравнения нескольких средних, что аналогично задаче однофакторного дисперсионного анализа.

Поэтому мы сосредоточимся на изложении пока малоизвестных и в силу этого обстоятельства редко применяемых методов двухфакторного непараметрического дисперсионного анализа для неполных данных. Для более глубокого изучения проблем неполных данных отсылаем заинтересованного читателя к сравнительно недавно переведенной на русский язык монографии [385].

5.1.3.1. Двухфакторный непараметрический дисперсионный анализ для неполных данных

Предположим, имеется таблица дисперсионного анализа с k уровнями фактора A $(i = 1, 2, \ldots, k)$ и m уровнями фактора B $(j = 1, 2, \ldots, m)$. Ранее мы рассматривали ситуацию, когда в каждой ячейке таблицы было одно наблюдение x_{ij}. Однако в практике часто могут быть случаи, когда часть требуемых значений x_{ij} может по тем или иным причинам отсутствовать: например, когда в эксперименте не все сочетания факторов A_i и B_j могут быть реализованы. Обычно при обработке неполных данных используется замена отсутствующих данных средними наблюдениями или средними рангами, вычисленными по некоторому множеству наблюдений. Однако это может приводить к ошибкам. В настоящем разделе приведены более точные критерии для такой ситуации.

5.1.3.1.1. Критерий Принтиса

Предположим, что двухфакторная таблица имеет m строк и k столбцов, и в каждой (i, j)-й ячейке находится либо одно наблюдение, либо ни одного. Пусть n_{ij} — число наблюдений в ячейке, расположенной на пересечении j-й строки и i-го столб-

ца ($i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, m$). Очевидно, что $n_{ij} = 1$ или $n_{ij} = 0$. Обозначим $n_j = \sum_{i=1}^{k} n_{ij}$. В каждой j-й строке ранжируем по возрастанию все n_j наблюдений и получаем последовательности рангов $R_{j1}, R_{j2}, \ldots, R_{jn_j}$ для всех $j = 1, 2, \ldots, m$.

Пусть S_i — множество строк (а, следовательно, и ранжировок), в которых есть наблюдение над i-м столбцом (т. е. ранг, относящийся к i-му столбцу). Далее вычисляем

$$a_i = \sum_{j \in S_i} \left(\frac{R_{ji}}{n_j + 1} - \frac{1}{2} \right); \quad w_{ii} = \frac{1}{12} \sum_{j \in S_i} \frac{n_j - 1}{n_j + 1}; \quad w_{ii'} = -\frac{1}{12} \sum_{j \in S_i \cap S_{i'}} \frac{1}{n_j + 1}$$

$$(i \neq i';\ i, i' = 1, \ldots, k).$$

Построим матрицу $W = \|w_{ii'}\|$, без ν-го столбца и ν-й строки, где $1 \leqslant \nu \leqslant k$ (т. е. матрица W строится вычеркиванием произвольно выбранных столбца и строки с одним и тем же номером из матрицы $\|w_{ii'}\|$). Строим вектор $a = (a_1, \ldots, a_{\nu-1}, a_{\nu+1} \ldots, a_k)$, т. е. удаляется компонента с номером ν вычеркнутых строки и столбца. Вычисляем матрицу W^{-1}, обратную к матрице W, и квадратичную форму $c = aW^{-1}a'$, где a' — вектор-столбец, получаемый транспонированием вектора-строки a.

При достаточно больших выборках ($m \geqslant 8$, $k \geqslant 6$) имеет место приближение [560, 561], из которого следует, что если $c \geqslant \chi^2_\alpha(k-1)$, то гипотеза об отсутствии влияния исследуемых факторов отклоняется с вероятностью α (здесь $\chi^2_\alpha(k-1)$ — α-квантиль распределения хи-квадрат с $f = k - 1$ степенями свободы).

При малых значениях m и k приближение действует плохо, однако таблицы критических значений этого критерия отсутствуют.

Задача 322. *Для данных, представленных в таблице, провести дисперсионный анализ критерием Принтиса при доверительной вероятности* $\alpha = 0{,}95$:

j	i		
	1	2	3
1	3,7	8,2	–
2	–	6,4	7,5
3	5,6	2,3	–

Имеем $m = k = 3$ и $n_j = 2$, $j = 1, 2, 3$. Производим ранжирование наблюдений по всем строкам и получаем таблицу рангов:

j	i		
	1	2	3
1	1	2	–
2	–	1	2
3	2	1	–

Множество строк S_1, в которых имеются наблюдения в первом ($i = 1$) столбце, будет $S_1 = \{1, 3\}$, и по аналогии $S_2 = \{1, 2, 3\}$, $S_3 = \{2\}$.

Далее вычисляем

$$a_1 = \sum_{j=1,3} \left(\frac{R_{j1}}{n_j + 1} - \frac{1}{2} \right) = \frac{R_{11}}{n_1 + 1} - \frac{1}{2} + \frac{R_{31}}{n_3 + 1} - \frac{1}{2} = \frac{1}{3} - \frac{1}{2} + \frac{2}{3} - \frac{1}{2} = 0;$$

$$a_2 = \sum_{j=1,2,3} \left(\frac{R_{j2}}{n_j + 1} - \frac{1}{2} \right) = \frac{2}{3} - \frac{1}{2} + \frac{1}{3} - \frac{1}{2} + \frac{1}{3} - \frac{1}{2} = -\frac{1}{6};$$

$$a_3 = \sum_{j=2} \left(\frac{R_{j3}}{n_j + 1} - \frac{1}{2} \right) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}.$$

Вычислим теперь диагональные элементы матрицы:

$$w_{11} = \frac{1}{12} \cdot \sum_{j=1,3} \frac{n_j - 1}{n_j + 1} = \frac{1}{12} \cdot \left(\frac{1}{3} + \frac{1}{3} \right) = \frac{2}{36};$$

$$w_{22} = \sum_{j=1,2,3} \frac{n_j - 1}{n_j + 1} = \frac{1}{12} \cdot \left(\frac{1}{3} + \frac{1}{3} + \frac{1}{3} \right) = \frac{3}{36};$$

$$w_{33} = \frac{1}{12} \cdot \sum_{j=3} \frac{n_j - 1}{n_j + 1} = \frac{1}{12} \cdot \frac{1}{3} = \frac{1}{36.}$$

Далее вычисляем остальные элементы матрицы:

$$w_{12} = w_{21} = -\frac{1}{12} \cdot \sum_{j \in S_1 \cap S_2} \frac{1}{n_j + 1} = -\frac{1}{12} \cdot \sum_{j=1,3} \frac{1}{n_j + 1} = -\frac{1}{12} \cdot \left(\frac{1}{3} + \frac{1}{3} \right) = -\frac{2}{36};$$

$$w_{13} = w_{31} = 0 \quad (S_1 \cap S_2 \text{ — пустое множество});$$

$$w_{23} = w_{32} = -\frac{1}{12} \cdot \sum_{j \in S_2 \cap S_3} \frac{1}{n_j + 1} = -\frac{1}{12} \cdot \sum_{j=2} \frac{1}{n_j + 1} = -\frac{1}{12} \cdot \frac{1}{3} = -\frac{1}{36}.$$

Таким образом, получаем матрицу

$$\|w_{ij}\| = \begin{Vmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{Vmatrix} = \begin{Vmatrix} \frac{2}{36} & -\frac{2}{36} & 0 \\ -\frac{2}{36} & \frac{3}{36} & -\frac{1}{36} \\ 0 & -\frac{1}{36} & \frac{1}{36} \end{Vmatrix}.$$

Отбросим 3-ю строку и 3-й столбец ($\nu = 3$), тогда имеем матрицу

$$W = \begin{Vmatrix} \frac{2}{36} & -\frac{2}{36} \\ -\frac{2}{36} & \frac{3}{36} \end{Vmatrix}.$$

Соответственно имеем вектор $a = \left(0, -\frac{1}{6}\right)$ (вместо $a = \left(0, -\frac{1}{6}, \frac{1}{6}\right)$ — так как вычеркивается третье значение).

Теперь вычисляем матрицу W^{-1}, обратную матрице W. Напомним, что обратная матрица W^{-1} связана с основной W соотношением $W \cdot W^{-1} = I$. Обратная матрица рассчитывается по формуле $W^{-1} = \frac{A}{\det W}$, где $\det W$ — определитель матрицы W, A — присоединенная матрица, являющаяся транспонированной матрицей алгебраических дополнений.

Алгебраическое дополнение A_{ij} элемента a_{ij} матрицы равно определителю матрицы, полученной вычеркиванием в первоначальной матрице i-й строки и j-го столбца, умноженному на $(-1)^{i+j}$.

Транспонирование матрицы достигается перестановкой ее строк и столбцов.

Для нашего случая

$$\det W = \det \begin{Vmatrix} \frac{2}{36} & -\frac{2}{36} \\ -\frac{2}{36} & \frac{3}{36} \end{Vmatrix} = \frac{2 \cdot 3}{36 \cdot 36} - \left(-\frac{2}{36} \right) \cdot \left(-\frac{2}{36} \right) = \frac{2}{36 \cdot 36}.$$

Для матрицы $W = \begin{Vmatrix} \frac{2}{36} & -\frac{2}{36} \\ -\frac{2}{36} & \frac{3}{36} \end{Vmatrix}$ матрица алгебраических дополнений имеет вид $\|A_{ij}\| = \begin{Vmatrix} \frac{3}{36} & \frac{2}{36} \\ \frac{2}{36} & \frac{2}{36} \end{Vmatrix}$. Присоединенная матрица получается транспонированием матрицы $\|A_{ij}\|$: $A = \begin{Vmatrix} \frac{3}{36} & \frac{2}{36} \\ \frac{2}{36} & \frac{2}{36} \end{Vmatrix}$. Окончательно получаем

$$W^{-1} = \frac{A}{\det W} = \frac{\begin{Vmatrix} \frac{3}{36} & \frac{2}{36} \\ \frac{2}{36} & \frac{2}{36} \end{Vmatrix}}{\frac{2}{36 \cdot 36}} = \begin{Vmatrix} \frac{3 \cdot 36 \cdot 36}{36 \cdot 2} & \frac{2 \cdot 36 \cdot 36}{36 \cdot 2} \\ \frac{2 \cdot 36 \cdot 36}{36 \cdot 2} & \frac{2 \cdot 36 \cdot 36}{36} \end{Vmatrix} = \begin{Vmatrix} 54 & 36 \\ 36 & 36 \end{Vmatrix}.$$

Далее вычисляем квадратичную форму ($\tilde{w}_{\imath\imath'}$ — элементы матрицы W^{-1})

$$c = \sum_{\imath=1}^{2} \sum_{\imath'=1}^{2} a_\imath \cdot a_{\imath'} \cdot \tilde{w}_{\imath\imath'} = 0 \cdot 0 \cdot 54 + \left(-\frac{1}{6}\right) \cdot 0 \cdot 36 + 0 \cdot \left(-\frac{1}{6}\right) \cdot 36 + \left(-\frac{1}{6}\right) \cdot \left(-\frac{1}{6}\right) \cdot 36 = 1.$$

Критическое значение из табл. 55 $\chi^2_{0,95}(2) = 5{,}99$.

Так как $c = 1 < \chi^2_{0,95}(2) = 5{,}99$, влияние факторов в таблице следует признать незначимым.

5.1.3.1.2. Критерий Мака–Скиллингса

Обозначения те же, что и в предыдущем разделе. В каждой ячейке допускается любое число наблюдений n_{ji} $(j = 1, \ldots, m;\ i = 1, \ldots, k)$; $N = \sum_{j,i} n_{ji}$. Процедура построения статистики критерия включает в себя следующие шаги. Все наблюдения j-й строки ранжируются по возрастанию от 1 до $n_j = \sum_{i=1}^{k} n_{ji}$. Обозначим через $r_{ji\nu}$ — ранг наблюдения $x_{ji\nu}$ в общей последовательности ($j = 1, \ldots, m$; $i = 1, \ldots, k$; $\nu = 1, \ldots, n_{ji}$).

Вычисляем

$$R_{ji} = \sum_{\nu=1}^{n_{\imath\jmath}} r_{ji\nu_{\jmath\imath}}; \quad R_i = \frac{1}{n_\jmath} \sum_{j=1}^{m} R_{ji} \quad (1 \leqslant j \leqslant m;\ 1 \leqslant i \leqslant k).$$

Будем рассматривать случай пропорциональных частот, т. е. когда $n_{ji} = \dfrac{N}{mk} = \dfrac{n_\imath \tilde{n}_\imath}{\sum_{\jmath,\imath} n_{ji}}$, где $\tilde{n}_i = \sum_{j=1}^{m} n_{ji}$. Для этого (достаточно распространенного) случая статистика критерия имеет вид [562]

$$T = \frac{12}{N\,(N+1)} \sum_{i=1}^{k} \tilde{n}_i \left(\tilde{R}_i - \frac{N+m}{2}\right)^2, \quad \text{где} \quad \tilde{R}_i = \frac{1}{n_{ji}} \sum_{j=1}^{m} R_{ji}.$$

При $N = \sum_{ji} n_{ji} \to \infty$ ($N > 15$) справедлива χ^2-аппроксимация. Влияние изучаемых факторов на поведение случайной величины с достоверностью α признается значимым, если $T \geqslant \chi^2_\alpha(k-1)$.

Критические значения $T_\alpha(m, k, n_{ji})$ статистики Мака–Скиллингса приведены в табл. 216.

Таблица 216

Критические значения T_α (m, k, n_{ji}) статистики Мака–Скиллингса [560, 562]

m	n_{ji}	Доверительная вероятность α							
		0,90				0,95			
		k				k			
		2	3	4	5	2	3	4	5
2	2	2,700	5,143	6,083	7,418	4,800	5,571	7,250	8,727
2	3	3,429	4,578	6,128	7,500	4,667	5,733	7,615	8,967
2	4	4,375	4,635	6,243	7,664	4,167	5,846	7,577	9,036
2	5	4,142	4,530	6,160	7,617	4,276	5,880	7,686	9,338
3	2	3,200	4,667	6,167	7,527	5,000	6,000	7,444	9,018
3	3	3,571	4,662	6,231	7,633	4,587	5,896	7,479	9,089
3	4	2,778	4,625	6,231	7,776	4,000	5,936	7,757	9,300
3	5	3,058	4,654	6,265	7,714	3,060	5,927	7,750	9,263
4	2	3,750	4,587	6,250	7,500	5,400	5,786	7,625	8,918
4	3	3,048	4,571	6,231	7,700	3,857	5,956	7,667	9,200
4	4	3,000	4,622	6,325	7,761	4,083	5,984	7,737	9,354
4	5	3,153	4,580	6,243	7,782	3,938	5,955	7,717	9,404
5	2	3,000	4,629	6,267	7,702	4,320	5,886	7,733	9,251
5	3	2,752	4,604	6,179	7,673	4,200	5,920	7,779	9,347
5	4	2,817	4,669	6,287	7,746	2,267	5,977	7,747	9,317
5	5	2,987	4,572	6,147	7,704	4,034	5,844	7,583	9,250

Задача 323. *Выполнить дисперсионный анализ данных, приведенных в таблице, критерием Мака–Скиллингса при доверительной вероятности* $\alpha = 0{,}95$:

j	i					
	1		2		3	
1	1,2	1,5	2,3	2,4	3,1	3,4
	1,3	1,7	2,5	2,6	3,5	3,3
2	1,4	1,7	3,7	3,1	2,9	2,1
	1,5	1,6	3,2	3,4	2,2	2,4
3	2,2	2,8	1,9	1,1	3,7	3,9
	2,1	2,4	1,6	1,4	3,1	3,3

Проведя ранжирование данных по строкам, получаем таблицу рангов $r_{ij\nu}$:

j	i					
	1		2		3	
1	1	2	5	6	9	12
	3	4	7	8	11	10
2	1	4	12	9	8	5
	2	3	10	11	6	7
3	6	8	4	1	11	12
	5	7	3	2	9	10

В нашем случае $n_{ji} = 4$; $m = 3$ и $k = 3$.

Суммы рангов $R_{ji} = \sum\limits_{\nu=1}^{4} r_{ji\nu}$ приведены в следующей таблице:

j	i		
	1	2	3
1	10	26	42
2	10	42	26
3	26	10	42

Имеем

$$N = \sum_{j,i} n_{ji} = 36; \quad n_{ji} = \frac{N}{mk} = 4,$$

т. е. имеет место вариант пропорциональных частот.

Далее находим

$$\tilde{R}_i = \frac{1}{n_{ji}} \cdot \sum_{j=1}^{3} R_{ji} = \frac{1}{4} \cdot \sum_{j=1}^{3} R_{ji}:$$

$$\tilde{R}_1 = \frac{10+10+26}{4} = 11{,}5; \quad \tilde{R}_2 = \frac{26+42+10}{4} = 19{,}5; \quad \tilde{R}_3 = \frac{42+26+42}{4} = 27{,}5.$$

Вычисляем статистику критерия

$$T = \frac{12}{N \cdot (N+m)} \cdot \sum_{i=1}^{3} \tilde{n}_i \cdot \left(\tilde{R}_i - \frac{N+m}{2}\right)^2 =$$

$$= \frac{12 \cdot 12}{36 \cdot (36+3)} \cdot \left[\left(11{,}5 - \frac{36+3}{2}\right)^2 + \ldots + \left(27{,}5 - \frac{36+3}{2}\right)^2\right] = 13{,}128.$$

Для $m = k = 3$, $n_{ji} = 4$ и $\alpha = 0{,}95$ из табл. 216 имеем $T_{0,95}(3,3,4) = 5{,}936$.

Так как $T = 13{,}128 > T_{0,95}(3,3,4) = 5{,}936$, следует признать влияние изучаемых факторов значимым.

5.1.3.1.3. Критерий Лемана–Мака

Обозначения аналогичны критерию Мака–Скиллингса. Ранжируем все $n_j = \sum\limits_{i=1}^{k} n_{ji}$ наблюдений внутри j-й строки и обозначим через $R_{ji\nu}$ ранг наблюдения $x_{ji\nu}$ $(\nu = 1, \ldots, n)$.

Вычисляем сумму и средний ранг наблюдения над j-м столбцом в i-й строке $R_{ji} = \frac{1}{n_{ji}} \sum\limits_{\nu=1}^{n_{ji}} R_{ji\nu}$, $j = 1, \ldots, m$; $i = 1, \ldots, k$.

Далее, вычисляем значение статистики Крускала–Уоллиса (см. раздел 4.2.1.2.1):

$$H_j = \frac{12}{n_j(n_j+1)} \sum_{i=1}^{k} n_{ji} \left(R_{ji} - \frac{n_j+1}{2}\right)^2 = \frac{12}{n_j(n_j+1)} \sum_{i=1}^{k} \frac{1}{n_{ji}} (R_{ji} n_{ji})^2 - 3(n_j+1) =$$

$$= \frac{12}{n_j(n_j+1)} \sum_{i=1}^{k} \frac{1}{n_{ji}} \left(\sum_{\nu=1}^{n_{ji}} R_{\nu_{ji\nu}}\right)^2 - 3(n_j+1).$$

Наконец вычисляем статистику критерия (сумму статистик Крускала–Уоллиса)

$$H = \sum_{j=1}^{m} H_j = \sum_{j=1}^{m} \frac{12}{n_j(n_j+1)} \sum_{i=1}^{k} n_{ji} \left(R_{ji} - \frac{n_j+1}{2}\right)^2.$$

Громоздкие таблицы критических значений статистики H приведены в [563]. Для частного случая $n_j = n$ (постоянное число наблюдений по строкам)

$$H_j^* = \frac{n(n+1)}{12} H_j; \quad H_j^{**} = \frac{n(n+1)}{12}(n - 1 - H_j).$$

Обозначим через l_j число пустых ячеек в j-й строке и введем величины

$$f_1 = \sum_{j=1}^{m} (l_j - 1) \quad \text{и} \quad f_2 = \sum_{j=1}^{m} (n - l_j).$$

Тогда статистика Лемана–Мака имеет вид

$$L = \frac{\sum_{j=1}^{m} H_j^* f_2}{f_1 \sum_{j=1}^{m} H_j^{**}}.$$

Ее критические значения равны

$$L_\alpha(f_1, f_2) = \frac{m(n-1) f_1 F_\alpha(f_1, f_2)}{f_2 + f_1 F_\alpha(f_1, f_2)},$$

где $F_\alpha(f_1, f_2)$ — α-квантиль F-распределения с f_1 и f_2 степенями свободы; для нахождения F_α рекомендуется пользоваться специальными таблицами ([24, 25, 29] и особенно [87]).

Влияние изучаемого фактора на случайные величины признается значимым с вероятностью α, если $L > L_\alpha(f_1, f_2)$.

Рассмотрим теперь случай, когда хотя бы два числа наблюдений в строках не равны: $n_{j_1} \neq n_{j_2}$; $j_1, j_2 \in \{1, \ldots, m\}$, $j_1 \neq j_2$. В этом случае гипотеза о влиянии факторов принимается, если $H \geqslant \chi_\alpha^2(f_1)$, где $\chi_\alpha^2(f_1)$ — α-квантиль распределения хи-квадрат с f_1 степенями свободы.

Задача 324. *Выполнить дисперсионный анализ данных, приведенных в задаче* 323, *критерием Лемана–Мака при доверительной вероятности* $\alpha = 0{,}95$.

Имеем $n_{ij} = 4$, $n_j = n_i = 12$, $m = 3$, $k = 3$. Суммы рангов по ячейкам таблицы приведены в таблице (см. задачу 323):

j	i		
	1	2	3
1	10	26	42
2	10	42	26
3	26	10	42

Вычисляем

$$H_1 = \frac{12}{n_1 \cdot (n_1 + 1)} \cdot \sum_{i=1}^{3} \frac{1}{n_{1i}} \cdot \left(\sum_{\nu=1}^{n_{1i}} R_{1i\nu} \right)^2 - 3 \cdot (n_1 + 1) =$$

$$= \frac{12}{12 \cdot 13} \cdot \frac{1}{4} \cdot \left(10^2 + 26^2 + 42^2\right) - 3 \cdot (12 + 1) = 9{,}8461.$$

Очевидно, что $H_1 = H_2 = H_3$ и $H = \sum_{j=1}^{3} H_j = 3 \cdot 9{,}8461 = 29{,}5383$.

Далее при $n_j = n = 12$ имеем

$$H_1^* = H_2^* = H_3^* = \frac{12 \cdot 13}{12} \cdot H_1 = 13 \cdot 9{,}8461 = 127{,}9993;$$

$$H_1^{**} = H_2^{**} = H_3^{**} = \frac{12 \cdot 13}{12} \cdot (12 - 1 - 9{,}8461) = 15{,}0007.$$

Число непустых ячеек в j-й строке равно $l_j = 3$ $(j = 1, 2, 3)$.
Находим

$$\sum_{j=1}^{3} H_j^{**} = 45{,}0021; \quad f_1 = \sum_{j=1}^{3} (l_j - 1) = 6; \quad f_2 = \sum_{j=1}^{3} (12 - l_j) = 27.$$

Окончательно имеем $L = \dfrac{383{,}9979}{6 \cdot 45{,}00021} = 38{,}398$. Из таблиц находим $F_{0,95}(6{,}27) = 2{,}459$ и $L_{0,95}(27) = \dfrac{3 \cdot (12 - 1) \cdot 6 \cdot 2{,}459}{27 + 6 \cdot 2{,}459} = 11{,}6$.

Так как $L = 38{,}398 > L_{0,95}(6{,}27) = 11{,}66$, влияние изучаемых факторов следует с достоверностью $\alpha = 0{,}95$ признать значимым.

5.2. Корреляционный анализ

Корреляционный анализ предполагает изучение зависимости между случайными величинами с одновременной количественной оценкой степени неслучайности их совместного изменения.

Изменение случайной величины y, соответствующее изменению случайной величины x, разбивается на две составляющие — стохастическую, связанную с неслучайной зависимостью y от x, и случайную (или статистическую), связанную со случайным характером поведения самих y и x.

Стохастическая составляющая связи между y и x характеризуется коэффициентом корреляции

$$\rho = \frac{\mathbf{M}\{[x-\mathbf{M}(x)][y-\mathbf{M}(y)]\}}{\sqrt{\mathbf{D}(x)\mathbf{D}(y)}},$$

где $\mathbf{M}(z)$ и $\mathbf{D}(z)$ — соответственно математическое ожидание и дисперсия случайной величины z.

Коэффициент корреляции показывает, насколько связь между случайными величинами близка к строго линейной. Если y и x распределены нормально, равенство $\rho = 0$ указывает на отсутствие линейной связи между ними. Значение $\rho = \pm 1$ соответствует строго линейной связи между y и x (знак указывает на направление связи).

Однако коэффициент корреляции ρ не учитывает возможной криволинейной связи между случайными величинами. Для учета таких связей используется корреляционное отношение, введенное К. Пирсоном.

Для двумерного ряда наблюдений, когда на каждом уровне одной переменной y_i наблюдаются n_i значений другой переменной x_{ij} $(j = 1, \ldots, n_i)$, корреляционное отношение определяется следующим образом

$$\eta_{xy}^2 = \frac{S_0^2}{S^2},$$

где S_0^2 — дисперсия рассеяния значений x_{ij}, связанная с влиянием группировки значений x_{ij} по i уровням переменной y; S^2 — дисперсия рассеяния значений x_{ij} без учета их группировки по уровням переменной y.

В нашем случае определено корреляционное отношение x по y. Перестановкой переменных по аналогии может быть определено η_{xy}^2 — корреляционное отношение y по x (тогда на каждом уровне переменной x_i наблюдается группа значений другой переменной y_{ij}). В общем случае $\eta_{xy}^2 \neq \eta_{yx}^2$.

Если y и x связаны строго линейно, то $\eta^2 = \rho^2 = 1$. Если между x и y существует линейная стохастическая связь, то $\rho^2 = \eta^2 < 1$. При нелинейной стохастической связи $\rho^2 < \eta^2 < 1$. В любом случае имеет место неравенство $0 \leqslant \rho^2 \leqslant \eta^2 \leqslant 1$ (равенство достигается только при строгой линейной связи между y и x).

5.2.1. Классический корреляционный анализ нормально распределенных случайных величин

5.2.1.1. Оценка коэффициента корреляции

Рассматриваются нормально распределенные случайные величины y и x — $(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_n, y_n)$. Выборочной оценкой коэффициента корре-

ляции ρ является случайная величина

$$r = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 \sum\limits_{i=1}^{n}(y_i - \bar{y})^2}},$$

где $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$; $\bar{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$; n — объем выборки.

При малых значениях n $(n < 15)$ лучшей оценкой коэффициента корреляции является

$$r^* = r\left[1 + \frac{1 - r^2}{2(n-3)}\right].$$

При $n > 200$ распределение выборочного коэффициента корреляции удовлетворительно аппроксимируется нормальным законом [1] со средним $\mathbf{M}(r)$ и дисперсией $\mathbf{D}(r)$:

$$\mathbf{M}(r) = \rho; \quad \mathbf{D}(r) = \frac{1 - \rho^2}{n - 1}.$$

При $n > 5$ распределение случайной величины [5]

$$z = \frac{1}{2}\ln\frac{1 - r}{1 + r} = \operatorname{arcth}(r)$$

удовлетворительно аппроксимируется нормальным распределением с параметрами

$$\mathbf{M}(z) = \frac{1}{2}\ln\left(\frac{1 - \rho}{1 + \rho}\right) = \operatorname{arcth}(\rho); \quad \mathbf{D}\,(z) = \frac{1}{n - 3}.$$

При $n > 10$ распределение случайной величины [564]

$$\tau = \frac{(r - \rho)\sqrt{n - 2}}{\sqrt{(1 - r^2)(1 - \rho^2)}}.$$

удовлетворительно аппроксимируется распределением Стьюдента с $f = n - 2$ степенями свободы.

Приведенные аппроксимации распределения выборочного коэффициента корреляции позволяют строить статистические критерии для проверки гипотез о существенности корреляционной связи и о возможных значениях коэффициента корреляции.

На практике наибольший интерес представляет задача проверки гипотезы о значимости корреляционной связи между случайными величинами, т. е. значимости отклонения коэффициента корреляции ρ от нуля. В принятых обозначениях проверяется нулевая гипотеза $H_0\colon |\rho| = 0$ против альтернативы $H_1\colon |\rho| \neq 0$.

Эта гипотеза проверяется сравнением выборочного значения коэффициента корреляции r с его критическим значением r_α, являющимся α-квантилью распределения r при $\rho = 0$. Корреляция между случайными величинами признается значимой, если $|r| \geqslant r_\alpha$. Критические значения r_α приведены в табл. 217.

Использование рассмотренных выше аппроксимаций приводит к следующим оценкам:

— при $n > 5$

$$r_\alpha = \frac{\exp\left(\frac{2}{\sqrt{n-3}}u_{\frac{1+\alpha}{2}}\right) - 1}{\exp\left(\frac{2}{\sqrt{n-3}}u_{\frac{1+\alpha}{2}}\right) + 1};$$

— при $n > 10$

$$r_\alpha = \sqrt{\frac{t^2_{\frac{1+\alpha}{2}}}{n - 2 + t^2_{\frac{1+\alpha}{2}}}};$$

— при $n > 200$

$$r_\alpha = \frac{1}{\sqrt{n-1}} u_{\frac{1+\alpha}{2}}.$$

Здесь u_α и t_α — α-квантили соответственно стандартного нормального распределения и распределения Стьюдента с $f = n - 2$ степенями свободы.

Таблица 217

Критические значения r_α выборочного коэффициента корреляции для $\rho = 0$ [25]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
3	0,988	0,997	1,000	13	0,476	0,553	0,684
4	0,900	0,950	0,990	14	0,457	0,532	0,661
5	0,805	0,878	0,959	15	0,441	0,514	0,641
6	0,729	0,811	0,917	16	0,426	0,497	0,623
7	0,669	0,754	0,874	17	0,412	0,482	0,606
8	0,621	0,707	0,834	18	0,400	0,468	0,590
9	0,582	0,666	0,798	19	0,389	0,456	0,575
10	0,549	0,632	0,765	20	0,378	0,444	0,561
11	0,521	0,602	0,735	21	0,369	0,433	0,549
12	0,497	0,576	0,708	22	0,360	0,423	0,537

Если гипотеза о значимости корреляции между случайными величинами не отклоняется, то можно построить доверительный интервал для истинного коэффициента корреляции по его выборочному значению. Впрочем, для корреляционного анализа это уже не столь важно, ибо его основная цель — установление значимости наблюдаемой связи.

Задача 325. *В результате наблюдений над случайными величинами x и y получена следующая совокупность данных ($n = 10$):*

x:	2	4	1	7	3	11	14	15	21	4
y:	7	6	4	11	2	21	31	23	40	15.

Необходимо проверить гипотезу о наличии корреляции между случайными величинами x и y с достоверностью $\alpha = 0{,}95$.

Находим

$$\bar{x} = \frac{1}{10} \cdot \sum_{i=1}^{10} x_i = 8{,}2; \quad \sum_{i=1}^{10} (x_i - \bar{x})^2 = 405{,}6; \quad \bar{y} = \frac{1}{10} \cdot \sum_{i=1}^{10} y_i = 16{,}0;$$

$$\sum_{i=1}^{10} (y_i - \bar{y})^2 = 1422; \quad \sum_{i=1}^{10} (x_i - \bar{x}) \cdot (y_i - \bar{y}) = 723.$$

Далее получаем оценки коэффициента корреляции

$$r = \frac{723}{\sqrt{405{,}6 \cdot 1422}} = 0{,}952; \quad r^* = 0{,}952 \cdot \left(1 + \frac{1 - 0{,}952^2}{2 \cdot 7}\right) = 0{,}958.$$

Из табл. 217 для $n = 10$ и $\alpha = 0{,}95$ находим $r_{0,95} = 0{,}632$.

Так как $r\ (r^*) = 0{,}952\ (0{,}958) > r_{0,95} = 0{,}632$, наличие зависимости между величинами x и y следует признать значимой с достоверностью $\alpha = 0{,}95$.

Если воспользоваться аппроксимациями (имея в виду, что $u_{\frac{1+0,95}{2}} = u_{0,975} = 1{,}96$), получим

$$r_{0,95} = \frac{\exp\left(\dfrac{2\cdot 1{,}96}{\sqrt{7}}\right) - 1}{\exp\left(\dfrac{2\cdot 1{,}96}{\sqrt{7}}\right) + 1} = 0{,}629,$$

что близко к точному значению $r_{0,95} = 0{,}632$.

С помощью t-приближения получим $(t_{\frac{1+\alpha}{2}}(f = n-2) = t_{0,975}(8) = 2{,}31)$

$$r_{0,95} = \left\{\frac{t^2_{0,975}(8)}{10 - 2 + t^2_{0,975}(8)}\right\}^{\frac{1}{2}} = \sqrt{\frac{2{,}31^2}{8 + 2{,}31^2}} = 0{,}632,$$

что совпадает с табличным значением.

Наконец, приближение для больших выборок дает

$$r_{0,95} = \frac{u_{0,975}}{\sqrt{n-1}} = \frac{1{,}96}{3} = 0{,}653.$$

5.2.1.2. Оценка корреляционного отношения

Предположим, что мы имеем n значений случайной величины y: $y_1, y_2, \ldots, y_k$. При $y = y_i$ наблюдаются n_i значений случайной величины x. Если $n = \sum_{i=1}^{k} n_i$; x_{ij} — j-е значение величины x, наблюдаемое при $y = y_i$ $(j = 1, 2, \ldots, n_i)$; $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$; $\bar{x} = \frac{1}{k}\sum_{i=1}^{k}\bar{x}_i$, то выборочная оценка корреляционного отношения x по y равна [132]

$$\eta^2_{xy} = \frac{\sum_{i=1}^{k} n_i\,(\bar{x}_i - \bar{x})^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2} = \frac{\sum_{i=1}^{k} n_i\bar{x}_i^2 - n\bar{x}^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - n\bar{x}^2}.$$

Проверка гипотезы $H_0\colon \eta^2 = 0$ против альтернативы $H_1\colon \eta^2 \neq 0$ производится с помощью статистики $l = \dfrac{\eta^2\,(n-k)}{(k-1)\left(1-\eta^2\right)}$.

Если $l \geqslant F_\alpha(f_1, f_2)$, то нулевая гипотеза отклоняется с достоверностью α. Здесь $F_\alpha(f_1, f_2)$ — α-квантиль F-распределения с $f_1 = k-1$ и $f_2 = n-k$ степенями свободы. При линейной связи между случайными величинами $\eta^2 = \rho^2$ и $\eta^2_{xy} = \eta^2_{yx}$. Следовательно, разность $\eta^2 - \rho^2$ может служить мерой нелинейности корреляционной связи. Проверка гипотезы $H_0\colon \eta^2 - \rho^2 = 0$ против альтернативы $H_1\colon \eta^2 - \rho^2 \neq 0$ может быть осуществлена с помощью статистики

$$l^* = \frac{\left(\eta^2 - \rho^2\right)(n-k)}{(k-2)\left(1-\eta^2\right)},$$

имеющей при справедливости нулевой гипотезы F-распределение с $f_1 = k - 2$ и $f_2 = n - k$ степенями свободы. Если $l^* \geqslant F_\alpha(f_1, f_2)$, то с вероятностью α гипотеза линейности корреляционной связи отклоняется. Следует помнить, что для оценки корреляционной связи x по y необходимо иметь несколько наблюдений x для различных y (и наоборот).

Задача 326. *Проверить линейность корреляционной связи для выборки*

y_i:	2	4	9	13	15
x_{ij}:	1, 3, 4	7, 8, 12	14, 19, 21	11, 9, 6	8, 7, 3

при доверительной вероятности $\alpha = 0{,}95$.

Имеем $k = 5$, $n_i = 3$ и $n = 15$. Вычисляем далее:

$$\bar{x}_1 = \frac{1+3+4}{3} = 2{,}66; \quad \bar{x}_2 = 9; \quad \bar{x}_3 = 18; \quad \bar{x}_4 = 8{,}67; \quad \bar{x}_5 = 6; \quad \sum_{i=1}^{5}\sum_{j=1}^{3} x_{ij}^2 = 1641;$$

$$\bar{x} = \frac{2{,}67 + 9 + 18 + 8{,}67 + 6}{5} = 8{,}864; \quad \sum_{i=1}^{5} n_i \cdot \bar{x}_i^2 = 3 \cdot \left(2{,}66^2 + 9^2 + \ldots + 6^2\right) = 1569{,}2136.$$

Тогда

$$\eta_{xy}^2 = \frac{1569{,}2136 - 15 \cdot 8{,}864^2}{1641 - 15 \cdot 8{,}864^2} = 0{,}8448.$$

Из таблиц находим $F_{0,95}(f_1,f_2) = F_{0,95}(5-1; 15-5) = F_{0,95}(4; 10) = 3{,}5$.

Вычисляем далее $l = \dfrac{\eta^2 \cdot (n-k)}{(k-1) \cdot \left(1-\eta^2\right)} = \dfrac{0{,}845 \cdot 10}{4 \cdot (1 - 0{,}845)} = 13{,}629$.

Полученная величина больше критического значения $F_{0,95}(4; 10) = 3{,}5$, следовательно, необходимо признать наличие существенной нелинейной связи между x и y.

Оценим теперь отклонение связи между x и y от линейной, для чего оценим коэффициент корреляции. Вместо значений x_{ij} на каждом уровне y_i будем использовать средние значения $\bar{x}_i$. Тогда ряд будет следующим:

$\bar{x}_i$:	2,66	9	18	8,66	6
y_i:	2	4	9	13	15.

Используя формулы из предыдущего раздела, получаем

$$\bar{x} = \frac{1}{5} \cdot \sum x_i = 8{,}864; \quad \bar{y} = \frac{1}{5} \cdot \sum y_i = 8{,}6; \quad \sum_{i=1}^{5} (\bar{x}_i - \bar{\bar{x}})^2 = 130{,}2187; \quad \sum_{i=1}^{5} (y_i - \bar{y}) = 125{,}5;$$

$$\sum_{i=1}^{5} (\bar{x}_i - \bar{\bar{x}}) \cdot (y_i - \bar{y}) = (-6{,}204) \cdot (-6{,}6) + \ldots + (-0{,}204) \cdot 6{,}4 = 24{,}775.$$

Тогда

$$r = \frac{24{,}775}{\sqrt{130{,}2187 \cdot 125{,}2}} = 0{,}194 \ (r^2 = 0{,}0376);$$

$$l^* = \frac{(0{,}8448 - 0{,}0376) \cdot (15-5)}{(5-2) \cdot (1 - 0{,}8448)} = 17{,}337.$$

Из таблиц имеем $F_{0,95}(5-2; 15-5) = F_{0,95}(3; 10) = 3{,}7$.

Так как $l^* = 17{,}337 > F_{0,95}(3; 10) = 3{,}7$, следует отклонить гипотезу о наличии линейной корреляционной связи между случайными величинами.

Отсюда следует поучительный вывод — незначимость коэффициента корреляции не означает отсутствия связи между исследуемыми величинами. Следует говорить об отсутствии линейной зависимости, так как незначимость коэффициента корреляции не исключает наличия нелинейной связи между случайными величинами.

5.2.1.3. Частная и множественная корреляции

При необходимости исследования связи между $\geqslant 3$ случайными величинами используются частные и множественные коэффициенты корреляции. Рассмотрим случай трех переменных — x, y и z (при числе переменных больше трех выражения для коэффициентов корреляции могут быть выписаны по аналогии).

Зависимость между двумя переменными x и y при фиксированной третьей переменной — z оценивается с помощью частного коэффициента корреляции $\rho_{xy,z}$. По аналогии можно определить частные коэффициенты корреляции по остальным парам переменных $\rho_{xz,y}$ и $\rho_{zy,x}$.

Выборочные частные (парные) коэффициенты корреляции определяются с помощью соотношений

$$r_{xy,z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}; \quad r_{xz,y} = \frac{r_{xz} - r_{xy}r_{zy}}{\sqrt{(1-r_{xy}^2)(1-r_{zy}^2)}}; \quad r_{zy,x} = \frac{r_{zy} - r_{zx}r_{yx}}{\sqrt{(1-r_{zx}^2)(1-r_{yx}^2)}};$$

$$r_{yz,x} = r_{ez,x}; \quad r_{xy,z} = r_{yx,z}; \quad r_{xz,y} = r_{zx,y}.$$

Так же, как и простые коэффициенты корреляции, парные коэффициенты принимают значения от -1 до $+1$. Гипотеза $H_0\colon \rho_{xy,z} = 0$ для коэффициента корреляции $\rho_{xy,z}$ (для остальных аналогично) проверяется с помощью статистики

$$t = \frac{\sqrt{n-k}\,r_{xy,z}}{\sqrt{1-r_{xy,z}^2}},$$

где k — число переменных (в нашем случае $k = 3$).

При справедливости H_0 величина t распределена в соответствии с распределением Стьюдента при $f = n - k$ степенях свободы.

При $|t| > t_{\frac{1+\alpha}{2}}(n-k)$ нулевая гипотеза H_0 отклоняется с вероятностью α. Множественная корреляция исследуется в случае, когда необходимо установить существенность взаимосвязи одной переменной с совокупностью остальных. Выборочные множественные коэффициенты корреляции обозначаются $r_{x,yz}$, $r_{y,xz}$, $r_{z,xy}$ и выражаются через парные коэффициенты корреляции с помощью соотношений

$$r_{x,yz}^2 = \frac{r_{xy}^2 + r_{xz}^2 - 2r_{xy}r_{zx}r_{yz}}{1-r_{yz}^2}; \quad r_{y,xz}^2 = \frac{r_{yx}^2 + r_{yz}^2 - 2r_{yx}r_{yz}r_{zx}}{1-r_{xz}^2};$$

$$r_{z,xy}^2 = \frac{r_{zx}^2 + r_{zy}^2 - 2r_{zx}r_{zy}r_{xy}}{1-r_{xy}^2}.$$

Между частными, множественными и обыкновенными парными коэффициентами корреляции имеют место так называемые контрольные соотношения:

$$r_{x,yz}^2 = 1 - \left(1-r_{xz}^2\right)\left(1-r_{xy,z}^2\right) = 1 - \left(1-r_{xy}^2\right)\left(1-r_{xz,y}^2\right);$$

$$r_{y,xz}^2 = 1 - \left(1-r_{yz}^2\right)\left(1-r_{yx,z}^2\right) = 1 - \left(1-r_{yx}^2\right)\left(1-r_{yz,x}^2\right);$$

$$r_{z,xy}^2 = 1 - \left(1-r_{zy}^2\right)\left(1-r_{zx,y}^2\right) = 1 - \left(1-r_{zx}^2\right)\left(1-r_{zy,x}^2\right).$$

Для проверки гипотезы $H_0\colon \rho_{x,yz} = 0$ используется статистика

$$F = \frac{r_{x,yz}^2}{1-r_{x,yz}^2}\,\frac{n-k}{k-1},$$

имеющая при справедливости H_0 F-распределение с $f_1 = k - 1$ и $f_2 = n - k$ степенями свободы (k — число переменных, в нашем случае $k = 3$).

Таблица 218

Критические значения $r_{1,23\ldots k}$ коэффициента множественной корреляции (k — число переменных, n — объем выборки)

$n-k$	Доверительная вероятность α							
	0,95				0,99			
	k				k			
	3	4	5	6	3	4	5	6
1	0,999	0,999	0,999	1,000	1,000	1,000	1,000	1,000
2	0,975	0,983	0,987	0,990	0,995	0,997	0,997	0,998
3	0,930	0,950	0,961	0,968	0,977	0,983	0,987	0,990
4	0,881	0,912	0,930	0,942	0,949	0,962	0,970	0,975
5	0,836	0,874	0,898	0,914	0,917	0,937	0,949	0,957
6	0,795	0,839	0,867	0,886	0,886	0,911	0,927	0,938
7	0,758	0,807	0,838	0,860	0,855	0,885	0,904	0,918
8	0,726	0,777	0,811	0,835	0,827	0,860	0,882	0,898
9	0,697	0,750	0,786	0,812	0,800	0,837	0,861	0,878
10	0,671	0,726	0,763	0,790	0,776	0,814	0,840	0,859
11	0,648	0,703	0,741	0,770	0,753	0,793	0,821	0,841
12	0,627	0,683	0,722	0,751	0,732	0,773	0,802	0,824
13	0,608	0,664	0,703	0,733	0,712	0,755	0,785	0,807
14	0,590	0,646	0,686	0,717	0,694	0,737	0,768	0,791
15	0,574	0,630	0,670	0,701	0,677	0,721	0,752	0,776
16	0,559	0,615	0,655	0,687	0,662	0,706	0,738	0,762
17	0,545	0,601	0,641	0,673	0,647	0,691	0,724	0,749
18	0,532	0,587	0,628	0,660	0,633	0,678	0,710	0,736
19	0,520	0,575	0,615	0,647	0,620	0,665	0,697	0,723
20	0,509	0,563	0,604	0,636	0,607	0,652	0,685	0,712
22	0,488	0,542	0,582	0,614	0,585	0,630	0,663	0,690
24	0,470	0,523	0,562	0,594	0,565	0,609	0,643	0,669
26	0,454	0,506	0,545	0,576	0,546	0,590	0,624	0,651
28	0,439	0,490	0,529	0,560	0,529	0,573	0,607	0,633
30	0,425	0,476	0,514	0,545	0,514	0,557	0,591	0,618
40	0,373	0,419	0,455	0,484	0,454	0,494	0,526	0,552
60	0,308	0,348	0,380	0,406	0,377	0,414	0,442	0,467

Если $F > F_\alpha(f_1, f_2)$, то соответствующая корреляция признается значимой. Критическое значение коэффициента множественной корреляции равно

$$r_{x.yz}(\alpha) = \sqrt{\frac{(k-1)\,F_\alpha(f_1,f_2)}{n-k+(k-1)\,F_\alpha(f_1,f_2)}}.$$

Корреляция признается значимой при $r_{x.yz} \geqslant r_{x.yz}(\alpha)$. Критические значения $r_{1.23\ldots k}$ (для общего случая k переменных) приведены в табл. 218.

Задача 327. *Вычислить коэффициенты частной и множественной корреляций и проверить их значимость при доверительной вероятности* $\alpha = 0{,}95$ *для данных, приведенных ниже* ($n = 10$, $k = 3$):

x_i:	1	3	4	7	12	4	19	21	1	3
y_i:	12	42	58	71	68	50	49	85	18	26
z_i:	41	12	7	3	14	27	38	13	64	75.

Найдем парные коэффициенты корреляции. Вычисляем коэффициент r_{xy}:

$$\sum_{i=1}^{10}(x_i-\bar{x})^2 = 484{,}5; \quad \sum_{i=1}^{10}(y_i-\bar{y})^2 = 6882{,}1; \quad \sum_{i=1}^{10}(x_i-\bar{x})\cdot(y_i-\bar{y}) = 1091;$$

$$\bar{x} = \frac{1}{10} \cdot \sum_{i=1}^{10} x_i = 7{,}5; \quad \bar{y} = \frac{1}{10} \cdot \sum_{i=1}^{10} y_i = 47{,}9; \quad r_{xy} = \frac{1091}{\sqrt{484{,}5 \cdot 6882{,}1}} = 0{,}597.$$

Вычисляем коэффициент r_{xz}:

$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 484{,}5; \quad \sum_{i=1}^{10} (z_i - \bar{z})^2 = 5498{,}4; \quad \sum_{i=1}^{10} (x_i - \bar{x})(z_i - \bar{z}) = -519;$$

$$\bar{x} = 7{,}5; \quad \bar{z} = 29{,}4; \quad r_{xz} = -\frac{519}{\sqrt{484{,}5 \cdot 5498{,}4}} = -0{,}318.$$

Вычисляем r_{yz}:

$$\sum_{i=1}^{10} (y_i - \bar{y})^2 = 6882{,}1; \quad \sum_{i=10}^{10} (z_i - \bar{z})^2 = 5498{,}4; \quad \sum_{i=1}^{10} (y_i - \bar{y}) \cdot (z_i - \tilde{z}) = -3172{,}66;$$

$$\bar{y} = 47{,}9; \quad \bar{z} = 29{,}4; \quad r_{yz} = -\frac{3172{,}66}{\sqrt{6862{,}1 \cdot 5498{,}4}} = -0{,}516.$$

Вычислим теперь частные коэффициенты корреляции:

$$r_{xy,z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{\left(1 - r_{xy}^2\right) \cdot \left(1 - r_{yz}^2\right)}} = \frac{0{,}597 - (-0{,}318) \cdot (-0{,}516)}{\sqrt{(1 - 0{,}318^2) \cdot (1 - 0{,}516^2)}} = 0{,}533;$$

$$r_{xz,y} = \frac{r_{xz} - r_{xy} \cdot r_{zy}}{\sqrt{\left(1 - r_{xy}^2\right) \cdot \left(1 - r_{zy}^2\right)}} = \frac{-0{,}318 - 0{,}597 \cdot (-0{,}516)}{\sqrt{(1 - 0{,}597^2) \cdot (1 - 0{,}516^2)}} = 0{,}014;$$

$$r_{zy,x} = \frac{r_{xy} - r_{zx} \cdot r_{yx}}{\sqrt{\left(1 - r_{zx}^2\right) \cdot \left(1 - r_{yx}^2\right)}} = \frac{-0{,}516 - (-0{,}318) \cdot 0{,}597}{\sqrt{(1 - 0{,}318^2) \cdot (1 - 0{,}597^2)}} = -0{,}429.$$

Вычислим множественные коэффициенты корреляции:

$$r_{x,yz}^2 = \frac{0{,}597^2 + 0{,}318^2 - 2 \cdot 0{,}597 \cdot (-0{,}18) \cdot (-0{,}516)}{1 - 0{,}516^2} = 0{,}356 \quad (r_{x,yz} = 0{,}597)\,;$$

$$r_{y,xz}^2 = \frac{0{,}597^2 + 0{,}516^2 - 2 \cdot 0{,}597 \cdot (-0{,}516) \cdot (-0{,}318)}{1 - 0{,}318^2} = 0{,}475 \quad (r_{y,xz} = 0{,}689)\,;$$

$$r_{z,xy}^2 = \frac{0{,}318^2 + 0{,}516^2 - 2 \cdot (-0{,}318) \cdot 0{,}597 \cdot (-0{,}516)}{1 - 0{,}597^2} = 0{,}266 \quad (r_{z,xy} = 0{,}516)\,.$$

Вычисляем t-статистики для проверки значимости частных коэффициентов корреляции

— для проверки $r_{xy,z}$: $t_{xy,z} = \dfrac{\sqrt{10-3} \cdot 0{,}533}{\sqrt{1 - 0{,}533^2}} = 1{,}667;$

— для проверки $r_{xz,y}$: $t_{xz,y} = \dfrac{\sqrt{7} \cdot (-0{,}014)}{\sqrt{1 - 0{,}014^2}} = -0{,}037;$

— для проверки $r_{zy,x}$: $t_{zy,x} = \dfrac{\sqrt{7} \cdot (-0{,}429)}{\sqrt{1 - 0{,}429^2}} = -1{,}256.$

Для $\alpha = 0{,}95$ и $f = n - k = 7$ из табл. 118 для t-распределения имеем $t_{\frac{1+0{,}95}{2}} = t_{0{,}975}(7) = 2{,}37$. Видим, что $|t_{xy,z}|\,, |t_{xz,y}|\,, |t_{zy,x}| < 2{,}37$.

Следовательно, наличие частной корреляции отклоняется с достоверностью $\alpha = 0{,}95$.

Для коэффициентов множественной корреляции находим критическое значение из табл. 218 при $k = 3$, $n - k = 7$ и $\alpha = 0{,}95$. Имеем $r_{1,23}(0{,}95) = 0{,}758$.

Так как ни один множественный коэффициент корреляции ($r_{x,yz} = 0{,}596$, $r_{y,xz} = 0{,}689$ и $r_{z,xy} = 0{,}516$) не превышает критическое значение $0{,}758$, то и наличие множественной корреляции отклоняется с достоверностью $0{,}95$.

В заключение проверим правильность вычислений, используя контрольные соотношения:

$$r_{x,yz}^2 = 0{,}596^2 = 0{,}356 = 1 - \left(1 - r_{xz}^2\right) \cdot \left(1 - r_{xy,z}^2\right) = 1 - \left(1 - 0{,}318^2\right) \cdot \left(1 - 0{,}533^2\right) = 0{,}356.$$

5.2.2. Непараметрический корреляционный анализ

Рассмотренные в разделе 5.2.1 методы корреляционного анализа предполагали нормальность распределения исследуемых величин. Для других распределений более эффективны методы изучения связи между случайными величинами, основанные на применении порядковых статистик, либо на замене наблюдаемых величин их рангами.

Такие методы, обладая повышенной устойчивостью к отклонениям распределения от нормального, в большинстве случаев позволяют упростить вычисления, оставляя на приемлемом уровне статистические характеристики получаемых заключений по гипотезам.

5.2.2.1. Оценивание корреляции с помощью порядковых статистик

5.2.2.1.1. Оценка корреляции с помощью тренда

Оценка наличия корреляции с помощью критериев тренда основано на следующей идее. Если значения одной переменной (например, x) предварительно упорядочить (например, по возрастанию), то поведение второй переменной (например, y) может служить индикатором искомой корреляции. В самом деле, наличие корреляции должно приводить к упорядочиванию значений второй переменной (т. е. к их тренду), отсутствие корреляции не должно изменять случайный характер поведения значений y при их размещении вдоль упорядоченной последовательности значений x.

Отсюда следует, что проверка ряда значений y на тренд любым из критериев тренда, изложенных в разделе 3.3, эквивалентно проверке наличия корреляции.

5.2.2.1.1.1. Критерий Кенуя

Рассмотрим некоторые из критериев, предложенных Кенуем [121] и предполагающих использование „быстрых“ критериев сравнения средних. Значения переменной x предварительно проранжируем по возрастанию: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ и зафиксируем ряд значений y_i, соответствующих x_i. Затем проверим наличие тренда в ряду значений y_i одним из „быстрых“ критериев Кенуя [121] или Кокс–Стюарта [565].

Для применения „быстрого“ критерия Кенуя поступаем следующим образом. Разбиваем проверяемую совокупность величин y_i объема n на k групп

$$y_{11}, y_{12}, \ldots, y_{1m}; \quad y_{21}, y_{22}, \ldots, y_{2m}; \quad \ldots; \quad y_{i1}, y_{i2}, \ldots, y_{im}; \quad \ldots;$$
$$y_{k1}, y_{k2}, \ldots, y_{km}, \quad (m = n/k)\,.$$

В каждой группе фиксируем крайние значения $y_{\min_i}$ и $y_{\max_i}$ для $i = 1, \ldots, k$. Затем отбираем значения $\max\limits_{1 \leqslant i \leqslant k} y_{\min_i}$ и $\min\limits_{1 \leqslant i \leqslant k} y_{\max_i}$. Вычисляем количество наблюдений n^-, для которых $y_i < \max\limits_{1 \leqslant i \leqslant k} y_{\min_i}$, и n^+, для которых $y_i > \min\limits_{1 \leqslant i \leqslant k} y_{\max_i}$.

Таблица 219

Критические значения n_α критерия Кенуя
(α — доверительная вероятность) [121]

α	k						
	2	3	4	5	6	8	10
0,95	8	17	27	37	47	70	93
0,99	12	22	33	45	57	83	110

Статистикой критерия является число $n = n^- + n^+$. Гипотеза о наличии корреляции принимается с достоверностью α, если $n > n_\alpha$. Критические значения n_α приведены в табл. 219.

5.2.2.1.1.2. Критерий Кокс–Стюарта

Рассмотрим теперь применение критерия Кокс–Стюарта для установления корреляции. Ряд наблюдений y_i, соответствующих упорядоченному ряду наблюдений x_i, разбивается на 3 приблизительно равные подвыборки (если $n/3$ — дробное число, то центральная подвыборка уменьшается (увеличивается) на 1). Сравниваем попарно $n' \approx n/3$ первых и $n' \approx n/3$ последних наблюдений. Если наблюдение из первой трети больше соответствующего наблюдения третьей трети, то поставим знак $+1$, в ином случае -1. Статистикой критерия является сумма T получаемых $+1$ и -1.

Корреляция между изучаемыми величинами с вероятностью α признается значимой, если $|T| > T_\alpha$. Значения T_α приведены в табл. 220.

Таблица 220

Критические значения T_α критерия Кокс–Стюарта
(α — доверительная вероятность)

n'	α		n'	α		n'	α		n'	α	
	0,95	0,99		0,95	0,99		0,95	0,99		0,95	0,99
6	6	–	12	8	10	18	10	12	24	12	14
7	7	–	13	9	11	19	11	13	25	11	15
8	8	8	14	10	12	20	10	14	26	12	14
9	7	9	15	9	11	21	11	13			
10	8	10	16	10	12	22	12	14			
11	9	11	17	9	13	23	11	15			

Эффективность этого критерия по сравнению с классическим корреляционным $\approx 50\%$, т. е. для обеспечения одинаковых статистических характеристик при проверке гипотез рассматриваемый критерий требует в 2 раза большего объема выборки.

Задача 328. *Для ряда пар случайных величин*

(x_i, y_i): (1, 51), (19, 52), (2, 48), (19, 14), (5, 51), (17, 14), (8, 49),
(10, 50), (13, 33), (51, 16), (20, 42), (22, 31), (48, 19), (47, 21),
(26, 41), (26, 53), (30, 58), (31, 43), (46, 21), (45, 1), (42, 11),
(33, 43), (33, 51), (35, 47), (36, 30), (40, 19), (41, 32), (44, 31)

установить наличие корреляции „быстрыми“ критериями Кенуя и Кокс–Стюарта.

Ранжируем величины x_i по возрастанию и получаем совместный ряд:

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
1	51	10	50	19	14	26	53	33	51	41	32	46	21
2	48	13	33	20	42	30	58	35	47	42	11	47	21
5	51	17	14	22	31	31	43	36	30	44	31	48	19
8	49	19	52	26	41	33	43	40	19	45	1	51	16

Используем критерий Кенуя. Разбиваем ряд y_i на 4 подвыборки:

y_1:	51	48	51	49	50	33	14
y_2:	52	14	42	31	41	53	58
y_3:	43	43	51	47	30	19	32
y_4:	11	31	1	21	21	19	16

Имеем

$$y_{\min_1} = 14; \quad y_{\max_1} = 51; \quad y_{\min_2} = 14; \quad y_{\max_2} = 58; \quad y_{\min_3} = 19; \quad y_{\max_3} = 51;$$
$$y_{\min_4} = 1; \quad y_{\max_4} = 31; \quad \max_{1 \leqslant i \leqslant 4} y_{\min_i} = 19; \quad \min_{1 \leqslant i \leqslant 4} y_{\max_i} = 31.$$

Далее сравнением устанавливаем количество значений, меньших, чем $\max y_{\min_i} = 19$ $(n^- = 5)$, и бо́льших, чем $\min y_{\max_i} = 31$ $(n^+ = 16)$.

Тогда статистика критерия равна $T = n^- + n^+ = 21$.

Из табл. 219 для $\alpha = 0{,}95$ и $k = 4$ находим $T_{0,95} = 27$.

Так как $T = 21 < T_{0,95} = 27$, корреляция с достоверностью $\alpha = 0{,}95$ признается незначимой.

Используем теперь критерий Кокс–Стюарта.

Имеем $n = 28$. Выделим $28/3 \approx 9$ первых наблюдений y_i и 9 последних наблюдений. Сравнивая их попарно, получаем последовательность

$$+1, \ +1, \ +1, \ +1, \ +1, \ +1, \ -1, \ +1, \ -1$$

и $T = 5$. Из табл. 220 находим для $n' = 9$ и $\alpha = 0{,}95$: $T_{0,95} = 7$.

Так как $|T| = 5 < T_{0,95} = 7$, корреляцию следует признать незначимой на уровне значимости 0,05.

5.2.2.1.2. Знаковый корреляционный критерий Нелсона

Критерий, предложенный Нелсоном [566], позволяет установить наличие корреляции, непрерывно анализируя совместное поведение пар (x_i, y_i) по мере их появления в эксперименте (в процессе). Критерий основан на числе знаков последовательного изменения величин пар (x_i, y_i).

Если $x_i > x_{i-1}$, $y_i > y_{i-1}$ или $x_i < x_{i-1}$, $y_i < y_{i-1}$, то паре (x_i, y_i) приписывается знак $+$, в ином случае знак $-$. Другими словами, если значения пар (x_i, y_i) изменились в одном направлении, то это отображается знаком $+$, в разных направлениях — знаком $-$. Если в паре одно или оба значения (x_i, y_i) не изменились, то этой паре приписывается значение 0. Статистикой критерия является наименьшее количество S знаков одного вида ($+$ или $-$). Корреляция признается значимой при $S > S_\alpha$ (S_α — критическое значение, приведенное в табл. 221, n — число анализируемых знаков).

Таблица 221

Критические значения S_α знакового критерия корреляции Нелсона (α — доверительная вероятность) [566]

n	α		n	α		n	α	
	0,95	0,99		0,95	0,99		0,95	0,99
11	2	1	24	7	6	37	12	11
12	2	2	25	7	7	38	13	12
13	3	2	26	8	7	39	13	12
14	3	2	27	8	7	40	14	13
15	3	3	28	9	8	50	18	17
16	4	3	29	9	8	55	20	19
17	4	3	30	9	9	60	22	21
18	5	4	31	10	9	65	25	23
19	5	4	32	10	9	70	27	25
20	5	5	33	11	10	75	29	28
21	6	5	34	11	10	80	31	30
22	6	5	35	12	11	85	34	32
23	7	6	36	12	11	90	36	34

При $n > 90$ сумма S распределена асимптотически нормально [566] и $S_\alpha \approx \frac{n}{2} + u_{1-\alpha}\sqrt{\frac{11n-2}{36}}$.

Эффективность этого метода $\approx 50\%$ от классического. Однако основные его достоинства — простота и возможность анализировать корреляцию непрерывно (по мере поступления данных) делает его весьма привлекательным для практического применения.

Задача 329. *Проверить гипотезу корреляции по данным задачи 328 критерием Нелсона при доверительной вероятности $\alpha = 0{,}95$.*

Введем данные анализа появления пар в таблицу:

i	x_i	y_i	Знак	i	x_i	y_i	Знак	i	x_i	y_i	Знак	i	x_i	y_i	Знак
1	1	51		8	10	50	+	15	26	41	–	22	33	43	–
2	19	52	+	9	13	33	–	16	26	53	0	23	33	51	0
3	2	48	+	10	51	16	–	17	30	58	+	24	35	47	–
4	19	14	–	11	20	42	–	18	31	43	–	25	36	30	–
5	5	51	–	12	22	31	–	19	46	21	–	26	40	19	–
6	17	14	–	13	48	19	–	20	45	1	+	27	41	32	+
7	8	49	–	14	47	21	–	21	42	11	–	28	44	31	–

Из таблицы видим, что наименьшее количество знаков одного вида (в нашем случае +) равно $S = 6$. Критическое значение из табл. 221 для $\alpha = 0{,}95$ и общего количества знаков любого вида (в нашем случае оно равно 25, в том числе 6 плюсов и 19 минусов) равно $S_{0,95} = 7$.

Так как $S = 6 < S_{0,95} = 7$, корреляция признается незначимой.

Для нормальной аппроксимации имеем

$$S_{0,95} = \frac{25}{2} + u_{0,05} \cdot \sqrt{\frac{11 \cdot 25 - 2}{36}} = \frac{25}{2} - 1{,}645 \cdot \sqrt{7{,}583} = 7{,}97,$$

что близко к табличному значению $S_{0,95} = 7$.

5.2.2.1.3. Квадрантный критерий

Рассматривается последовательность случайных величин x и y с выборочными медианами $\tilde{x}$ и $\tilde{y}$. Введем обозначения

$$S_1 = \begin{cases} 1, & \text{если } x_i > \tilde{x} \text{ и } y_i > \tilde{y}; \\ 0 & \text{в остальных случаях;} \end{cases} \qquad S_2 = \begin{cases} 1/2, & \text{если } x_i = \tilde{x} \text{ и } y_i > \tilde{y}; \\ 0 & \text{в остальных случаях;} \end{cases}$$

$$S_3 = \begin{cases} \frac{1}{2}, & \text{если } x_i > \tilde{x} \text{ и } y_i = \tilde{y}; \\ 0 & \text{в остальных случаях;} \end{cases} \qquad S_4 = \begin{cases} \frac{1}{4}, & \text{если } x_i = \tilde{x} \text{ и } y_i = \tilde{y}; \\ 0 & \text{в остальных случаях.} \end{cases}$$

Статистика $S = S_1 + S_2 + S_3 + S_4$ может быть использована для оценки корреляции между случайными величинами x и y [365] (при n четном очевидно, что $S = S_1$). Критерий называется квадрантным, так как статистика S основана на числе наблюдений в квадрантах, на которые плоскость xy делится прямыми $x = \tilde{x}$ и $y = \tilde{y}$.

Гипотеза о наличии корреляции отклоняется, если $S_1(\alpha) < S < S_2(\alpha)$ (критические значения $S_1(\alpha)$ и $S_2(\alpha)$ приведены в табл. 222).

Таблица 222

Критические значения $S_1(\alpha)$ и $S_2(\alpha)$ квадрантного критерия корреляции (α — доверительная вероятность) [121]

n	α				n	α			
	0,95		0,99			0,95		0,99	
	S_1	S_2	S_1	S_2		S_1	S_2	S_1	S_2
$8 \div 9$	0	4	–	–	$74 \div 74$	13	24	12	25
$10 \div 11$	0	5	0	5	$76 \div 77$	14	24	12	26
$12 \div 13$	0	6	0	6	$78 \div 79$	14	25	13	26
$14 \div 15$	1	6	0	7	$80 \div 81$	15	25	13	27
$16 \div 17$	1	7	0	8	$82 \div 83$	15	26	14	27
$18 \div 19$	1	8	1	8	$84 \div 85$	16	26	14	28
$20 \div 21$	2	8	1	9	$86 \div 87$	16	27	15	28
$22 \div 23$	2	9	2	9	$88 \div 89$	16	28	15	29
$24 \div 25$	3	9	2	10	$90 \div 91$	17	28	15	30
$26 \div 27$	3	10	2	11	$92 \div 93$	17	29	16	30
$28 \div 29$	3	11	3	11	$94 \div 95$	18	29	16	31
$30 \div 31$	4	11	3	12	$96 \div 97$	18	30	17	31
$32 \div 33$	4	12	3	13	$98 \div 99$	19	30	17	32
$34 \div 35$	5	12	4	13	$100 \div 101$	19	31	18	32
$36 \div 37$	5	13	4	14	$110 \div 111$	21	34	20	35
$38 \div 39$	6	13	5	14	$120 \div 121$	24	36	22	38
$40 \div 41$	6	14	5	15	$130 \div 131$	26	39	24	41
$42 \div 43$	6	15	5	16	$140 \div 141$	28	42	26	44
$44 \div 45$	7	15	6	16	$150 \div 151$	31	44	29	46
$46 \div 47$	7	16	6	17	$160 \div 161$	33	47	31	49
$48 \div 49$	8	16	7	17	$170 \div 171$	35	50	33	52
$50 \div 51$	8	17	7	18	$180 \div 181$	37	53	35	55
$52 \div 53$	8	18	7	19	$200 \div 201$	42	58	40	60
$54 \div 55$	9	18	8	19	$220 \div 221$	47	63	44	66
$56 \div 57$	9	19	8	20	$240 \div 241$	51	69	49	71
$58 \div 59$	10	19	9	20	$260 \div 261$	56	74	54	76
$60 \div 61$	10	20	9	21	$280 \div 281$	61	79	58	82
$62 \div 63$	11	20	9	22	$300 \div 301$	66	84	63	87
$64 \div 65$	11	21	10	22	$320 \div 321$	70	90	67	93
$66 \div 67$	12	21	10	23	$340 \div 341$	75	95	72	98
$68 \div 69$	12	22	11	23	$360 \div 361$	80	100	77	103
$70 \div 71$	12	23	11	24	$380 \div 381$	84	106	81	109
$72 \div 73$	13	23	12	24	400	89	111	86	114

При $n > 100$ может быть использована аппроксимация

$$S(\alpha) = \begin{cases} \dfrac{n}{4}\left(1 + \dfrac{1}{\sqrt{n-1}}\, u_{\frac{1+\alpha}{2}}\right) & \text{при } n = 2k; \\ \dfrac{n}{4}\left(1 + \dfrac{n}{\sqrt{n-1}}\, u_{\frac{1+\alpha}{2}}\right) & \text{при } n = 2k-1. \end{cases}$$

Гипотеза наличия корреляции в этом случае отклоняется с достоверностью α, если $|S| < S(\alpha)$ (u_γ — γ-квантиль стандартного нормального распределения). Критерий обладает эффективностью $\approx 41\%$ от классического корреляционного критерия, однако он очень прост для вычислений.

Напомним, что медиана упорядоченного ряда $z_1 \leqslant z_2 \leqslant \ldots \leqslant z_n$ определяется соотношением

$$\tilde{z} = \begin{cases} z_{\frac{n+1}{2}} & \text{при } n = 2k-1; \\ \frac{1}{2}\left(z_{\frac{n}{2}} + z_{\frac{n+2}{2}}\right) & \text{при } n = 2k. \end{cases}$$

Несколько иная версия квадрантного критерия рассмотрена в [567, 568] — критерий Эландта. Его статистикой при n четном является

$$U = \sum_{i=1}^{n} U_i, \quad \text{где} \quad U_i = \begin{cases} 1 & \text{при } (x_i - \tilde{x})(y_i - \tilde{y}) > 0; \\ 0 & \text{при } (x_i - \tilde{x})(y_i - \tilde{y}) \leqslant 0. \end{cases}$$

При справедливости гипотезы о наличии корреляции

$$\mathbf{P}(U \geqslant U^*) = \frac{1}{C_n^{\frac{n}{2}}} \sum_{i=\frac{U^*}{2}}^{\frac{n}{2}} \left(C_{\frac{n}{2}}^i\right)^2.$$

Если $\mathbf{P}(U \geqslant U^*) > \alpha$, то с достоверностью α наличие корреляции отклоняется (U^* — выборочное значение статистики U).

Задача 330. *Проверить гипотезу корреляции для данных задачи 328 квадрантным критерием при доверительной вероятности* $\alpha = 0{,}95$.

Для упорядоченного ряда значений

x_i:	1	2	5	8	10	13	17	19	19	20	22	26	26	30
	31	33	33	35	36	40	41	42	44	45	46	47	48	51

имеем медиану $\tilde{x} = \frac{1}{2} \cdot (x_{14} + x_{15}) = \frac{30+31}{2} = 30{,}5$.

Для упорядоченного ряда

y_i:	1	11	14	16	19	19	21	21	30	31	32	33
	41	42	43	43	48	49	50	51	51	52	53	58

имеем $\tilde{y} = \frac{y_{14} + y_{15}}{2} = \frac{33+41}{2} = 37$. Далее находим количества пар, для которых $x_i > \tilde{x}$ и $y_i > \tilde{y}$; $x_i = \tilde{x}$ и $y_i > \tilde{y}$; $x_i > \tilde{x}$ и $y_i = \tilde{y}$; $x_i = \tilde{x}$ и $y_i = \tilde{y}$: соответственно $S_1 = 4$, $S_2 = 0$, $S_3 = 0$, $S_4 = 0$. Тогда $S = S_1 = 4$. Из табл. 222 для $n = 28$ и $\alpha = 0{,}95$ имеем $S_1(0,95) = 3$ и $S_2(0,95) = 11$.

Так как $S_1(0{,}95) = 3 < S = 4 < S_2(0{,}95) = 11$, с достоверностью $\alpha = 0{,}95$ наличие корреляции отклоняется. Используем теперь аппроксимацию

$$S\,(0{,}95) = \frac{28}{4} \cdot \left(1 + \frac{1}{\sqrt{27}} \cdot u_{0,975}\right) = 7 \cdot \left(1 + \frac{1{,}96}{\sqrt{27}}\right) = 9{,}6.$$

Так как $S = 4 < S(0{,}95) = 9{,}6$, то и в этом случае наличие корреляции отклоняется.

Применим теперь критерий Эландта. Имеем последовательность значений

U_i: 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.

Тогда получаем $U^* = \sum U_i = 4$. Вычисляем далее

$$\mathbf{P}(U \geqslant 4) = \frac{1}{C_{28}^{14}} \cdot \sum_{i=2}^{14} \left(C_{14}^i\right)^2 = \frac{\left(C_{14}^2\right)^2 + \left(C_{14}^3\right)^2 + \ldots + \left(C_{14}^{14}\right)^2}{C_{28}^{14}} =$$

$$= \frac{1}{40116600} \cdot \left(91^2 + 364^2 + 1001^2 + \ldots + 91^2 + 14^2 + 1^2\right) = 0{,}999.$$

Так как $\mathbf{P}(U \geqslant 4) = 0{,}999 > \alpha = 0{,}95$, гипотеза корреляции отклоняется.

5.2.2.1.4. Угловой критерий Олмстеда–Тьюки

Сущность критерия сводится к следующему. Двумерная диаграмма (x, y) сначала делится вертикальной прямой $x = \tilde{x}$ на две части и горизонтальной прямой $y = \tilde{y}$ на две части ($\tilde{x}, \tilde{y}$ — медианы).

Если n — нечетное число, то медианы проходят через одну из точек (x_i, y_i), которую следует исключить из рассмотрения. В результате получаем четыре квадранта. Квадрантам, для которых $(x_i - \tilde{x})(y_i - \tilde{y}) > 0$ (правый верхний и левый нижний), приписываем знак $+$, а квадрантам, для которых $(x_i - \tilde{x})(y_i - \tilde{y}) < 0$ (левый верхний и правый нижний) — знак $-$.

Затем, двигаясь слева направо (от $x_{\min}$ к $x_{\max}$), подсчитываем количество встречающихся подряд точек a_1, для которых $y_i > \tilde{y}$ ($y_i < \tilde{y}$). По аналогии, двигаясь справа налево (от $x_{\max}$ к $x_{\min}$), подсчитываем количество встречающихся подряд точек a_2, для которых $y_i < \tilde{y}$ ($y_i > \tilde{y}$). Затем, двигаясь сверху вниз (от $y_{\max}$ к $y_{\min}$), подсчитываем количество встречающихся подряд точек a_3, для которых $x_i < \tilde{x}$ ($x_i > \tilde{x}$), и, наконец, двигаясь снизу вверх (от $y_{\min}$ к $y_{\max}$), подсчитываем количество встречающихся подряд точек a_4, для которых $x_i > \tilde{x}$ ($x_i < \tilde{x}$). Знаки значениям a_1, a_2, a_3, a_4 присваиваются в зависимости от квадранта, в котором располагаются отобранные точки.

Статистикой критерия является абсолютная сумма

$$Q = \left| \sum_{i=1}^{4} a_i \right|,$$

критические значения которой приведены в табл. 223.

Таблица 223

Критические значения Q_α угловой статистики Олмстеда–Тьюки

(α — доверительная вероятность) [24]

α	0,90	0,95	0,99	0,999
Q_α	9	11	$14 \div 15$	$18 \div 21$

Меньшая величина Q_α применяется для выборок большого объема, бóльшая величина — для выборок малого объема.

При $Q > Q_\alpha$ наличие корреляции признается значимым с вероятностью α. Если $Q > 2n$ при $n \leqslant 6$, то критерием Олмстеда–Тьюки пользоваться не рекомендуется.

Критерий обладает низкой эффективностью ($\approx 25\%$ по сравнению с классическими оценками) и рекомендуется к применению при больших объемах выборок ($n > 50$). При $n > 50$ вероятность того, что $Q > Q_\alpha$, оценивается по формуле $\alpha = 1 - \dfrac{9Q^3 + 9Q^2 + 168Q + 208}{2162 \cdot 2^Q}$.

Задача 331. *Проверить наличие корреляции между исследуемыми данными критерием Олмстеда–Тьюки при доверительной вероятности* $\alpha = 0{,}95$ *в условиях задачи* 328.

В нашем случае $\tilde{x} = 30{,}5$ и $\tilde{y} = 37$. Движемся слева направо (от $x_{\min} = 1$ к $x_{\max} = 51$) и подсчитываем количество встречающихся подряд точек (x_i, y_i), для которых $y_i < \tilde{y}$. Видим, что таких точек $a_1 = 5$ — это точки $(1, 51)$, $(2, 48)$, $(5, 51)$, $(8, 49)$, $(10, 50)$. Все эти точки расположены в левом верхнем квадранте (так как $x_i < \tilde{x}$ и $y_i > \tilde{y}$), значение a_1 должно учитываться со знаком $-$.

Движемся теперь справа налево (от $x_{\max} = 51$ к $x_{\min} = 1$) и подсчитываем количество встречающихся подряд точек, для которых $y_i < \tilde{y}$, — их количество $a_2 = 10$ — это точки $(51, 16)$, $(48, 19)$, $(47, 21)$, $(46, 21)$, $(45, 1)$, $(44, 31)$, $(42, 11)$, $(41, 32)$, $(40, 19)$, $(36, 30)$.

Так как все эти точки расположены в правом нижнем квадранте ($x_i > \tilde{x}$ и $y_i < \tilde{y}$), то сумма $a_2 = 10$ должна учитываться со знаком $-$.

Далее движемся сверху вниз (от $y_{\max} = 58$ к $y_{\min} = 1$) и подсчитываем количество встречающихся подряд точек a_3, для которых $x_i < \tilde{x}$: это точки $(30, 58)$, $(26, 53)$, $(19, 52)$, $(1, 51)$. Так как эти точки находятся в левом верхнем квадранте, то значение $a_3 = 4$ учитывается со знаком $-$.

И, наконец, движемся снизу вверх (от $y_{\min} = 1$ к $y_{\max} = 58$) и подсчитываем количество встречающихся подряд точек, для которых $x_i > \tilde{x}$, их $a_4 = 2$. Это точки $(45, 1)$, $(42, 11)$. Все они располагаются в правом нижнем квадранте, и поэтому $a_4 = 2$ также учитывается со знаком $-$.

Окончательно имеем $Q = |a_1 + a_2 + a_3 + a_4| = |-5 - 10 - 4 - 2| = 21$.

Легко видеть, что критерий приводит к принятию гипотезы о наличии корреляции, так как $Q = 21 > Q_{0,95} = 11$. Однако не следует этот вывод считать достоверным, так как для принятия решения критерием Олмстеда–Тьюки рекомендуется использовать выборки объема $n > 50 \div 100$ в силу малой эффективности критерия. Здесь мы рассмотрели пример только для демонстрации вычислительной техники критерия.

Теперь вычислим вероятность

$$\alpha = 1 - \frac{9 \cdot 21^3 + 9 \cdot 21^2 + 168 \cdot 21 + 208}{216 \cdot 2^{21}} = 0{,}9998.$$

Так как $\alpha = 0{,}9998 > 0{,}95$, гипотеза корреляции принимается (комментарий см. выше).

5.2.2.1.5. Приближенный критерий Шахани

В совокупности значений пар (x_i, y_i) выделим порядковые статистики $x_{[0,3n]}, x_{[0,7n]}, y_{[0,3n]}, y_{[0,7n]}$, т. е. значения x и y, которые в упорядоченных по возрастанию рядах x_i и y_i занимают места с номерами $[0{,}3n]$ и $[0{,}7n]$.

Далее обозначим количество наблюдений, попавших в угол, для которого $x_i < x_{[0,3n]}$, $y > y_{[0,7n]}$, через a_1; количество наблюдений, попавших в угол $x_i > x_{[0,7n]}$, $y > y_{[0,7n]}$, через a_2; количество наблюдений, попавших в угол $x_i > x_{[0,7n]}$, $y_i < y_{[0,3n]}$, через a_3; количество наблюдений в угле $x_i < x_{[0,3n]}$, $y_i < y_{[0,3n]}$ через a_4. Статистика критерия равна $q = |a_1 + a_3 - a_2 - a_4|$. Критическое значение q-статистики равно $q_\alpha = \left\{ \sum_{i=1}^{4} a_i \right\}^{\frac{1}{2}} u_{\frac{1+\alpha}{2}}$, где u_γ — γ-квантиль стандартного нормального распределения.

При $q > q_\alpha$ корреляция между исследуемыми величинами признается значимой. Эффективность критерия по сравнению с классическим коэффициентом корреляции $\approx 0{,}67$ [121].

Задача 332. *Проверить гипотезу корреляции для данных задачи* 328 *критерием Шахани.*

Имеем $x_{[0,3n]} = x_{[0,3 \cdot 28]} = x_9 = 19$ ($[a]$ — ближайшее большее целое к a). Далее $x_{[0,7 \cdot n]} = x_{20} = 40$; $y_{[0,3 \cdot n]} = y_9 = 21$; $y_{[0,7n]} = y_{20} = 48$.

Имеем в угле $x_i < 19$, $y_i > 48$: $a_1 = 5$; в угле $x_i > 40$, $y_i > 48$: $a_2 = 0$; в угле $x_i > 40$, $y_i < 21$: $a_3 = 4$; в угле $x_i < 19$, $y_i < 21$: $a_4 = 1$ точку. Тогда

$$q = |5 + 4 - 0 - 1| = 8; \quad q_{0,95} = \sqrt{5 + 4 + 0 + 1}\, u_{0,975} = 3{,}16 \cdot 1{,}96 = 6{,}2.$$

Так как $q = 8 > q_{0,95} = 6{,}2$ корреляция признается значимой.

5.2.2.1.6. Сериальный критерий Шведа–Эйзенхарта

Совокупность n пар (x_i, y_i) разбивается на две равные совокупности, отвечающие условиям $y_i > \tilde{y}$ и $y_i < \tilde{y}$ ($\tilde{y}$ — медиана ряда y_i, при n нечетном значение $y_i = \tilde{y}$ исключается). Затем наблюдения ранжируются по возрастающим значениям x_i.

Для последовательных пар значений (x_i, y_i) с $y_i > \tilde{y}$ будем применять символ a, для последовательных пар (x_i, y_i) с $y_i < \tilde{y}$ — символ b. В результате получим последовательность элементов вида $a, b, b, a, a, a, b, \ldots$. Последовательность элементов одного вида, ограниченная с двух сторон элементами другого вида (замыкающие интервал последовательности одного вида ограничены с одной стороны последовательностями другого вида), называется серией. Количество m серий является статистикой рассматриваемого критерия [121].

Корреляция признается значимой, если $m \leqslant m_\alpha$ (критические значения m_α приведены в табл. 224).

Таблица 224

Критические значения m_α критерия Шведа–Эйзенхарта
(α — доверительная вероятность) [121]

n	α		n	α		n	α	
	0,95	0,99		0,95	0,99		0,95	0,99
$8 \div 9$	2	–	$20 \div 21$	6	5	$32 \div 33$	11	10
$10 \div 11$	3	2	$22 \div 23$	7	6	$34 \div 35$	12	10
$12 \div 13$	3	2	$24 \div 25$	8	7	$36 \div 37$	13	11
$14 \div 15$	4	3	$26 \div 27$	9	7	$38 \div 39$	14	12
$16 \div 17$	5	4	$28 \div 29$	10	8	$40 \div 41$	15	13
$18 \div 19$	6	4	$30 \div 31$	11	9			

При четных $n > 40$ можно использовать приближения

$$m_{0,95} = \left[\frac{n+1}{2} - 0{,}82\sqrt{n-1}\right]; \quad m_{0,99} = \left[\frac{n+1}{2} - 1{,}16\sqrt{n-1}\right],$$

где $[z]$ — ближайшее меньшее целое к z.

Задача 333. *Проверить гипотезу корреляции для данных задачи 328 критерием Шведа–Эйзенхарта.*

В нашем случае $\tilde{y} = 37$. Будем обозначать пары (x_i, y_i), в которых $y_i > \tilde{y} = 37$, символом a, а пары, в которых $y_i < \tilde{y} = 37$, символом b. Располагая пары в порядке увеличения значений x_i,получаем последовательность

$$\underline{aaaaa},\ \underline{bb},\ \underline{a},\ \underline{b},\ \underline{a},\ \underline{b},\ \underline{aaaaaaa},\ \underline{bbbbbbbb}.$$

Видим, что в полученной последовательности содержится $m = 8$ серий (4 серии элементов a и 4 серии элементов b). Из табл. 224 для $n = 28$ и $\alpha = 0{,}95$ находим $m_{0,95} = 10$. Так как $m = 8 < m_{0,95} = 10$, корреляция признается значимой. Приближение $m_{0,95} = \left[\frac{28+1}{2} - 0{,}82 \cdot \sqrt{28-1}\right] = 10$ дает такой же результат.

5.2.2.1.7. Критерий автокорреляции Кенуя

Критерий позволяет установить наличие корреляции в ряду пар значений (x_i, y_i), расположенных по возрастанию одной из величин (например, x). Под автокорреляцией понимается наличие зависимости значений переменной величины от порядкового номера ее расположения в ряду данных. Проверке такой зависимости и служит критерий Кенуя.

Критерий строится следующим образом. Все $(n-1)$ пар значений располагаются в порядке возрастания x_i от $x_{\min}$ до $x_{\max}$ и разбиваются на две группы, с $y_i > \tilde{y}$ и $y_i < \tilde{y}$ ($\tilde{y}$ — медиана). Затем последовательно рассматриваем пары, для которых справедливо

$$(x_i, y_i > \tilde{y}), (x_{i+1}, y_{i+1} > \tilde{y}) \quad \text{или} \quad (x_i, y_i < \tilde{y}), (x_{i+1}, y_{i+1} < \tilde{y}).$$

Другими словами, определяется количество последовательных пар точек, находящихся по какую-либо одну сторону от медианы. Количество таких пар N является статистикой критерия. Если $N > n_\alpha$, то корреляция признается значимой. Критические значения n_α приведены в табл. 225.

Таблица 225

Критические значения N_α критерия автокорреляции Кенуя
(α — доверительная вероятность) [121]

n	α		n	α		n	α	
	0,95	0,99		0,95	0,99		0,95	0,99
$8 \div 9$	6	–	$40 - 41$	25	27	$72 - 73$	43	46
$10 \div 11$	7	8	$42 \div 43$	26	28	$74 \div 75$	44	47
$12 \div 13$	9	10	$44 \div 45$	27	30	$76 \div 77$	45	48
$14 \div 15$	10	11	$46 \div 47$	29	31	$78 \div 79$	46	49
$16 \div 17$	11	12	$48 \div 49$	30	32	$80 \div 81$	47	50
$18 \div 19$	12	14	$50 \div 51$	31	33	$82 \div 83$	48	51
$20 \div 21$	14	15	$52 \div 53$	32	34	$84 \div 85$	49	53
$22 \div 23$	15	16	$54 \div 55$	33	35	$86 \div 87$	51	54
$24 \div 25$	16	17	$56 \div 57$	34	37	$88 \div 89$	52	55
$26 \div 27$	17	19	$58 \div 59$	35	38	$90 \div 91$	53	56
$28 \div 29$	18	20	$60 \div 61$	36	39	$92 \div 93$	54	57
$30 \div 31$	19	21	$62 \div 63$	37	40	$94 \div 95$	55	58
$32 \div 33$	21	22	$64 \div 65$	39	41	$96 \div 97$	56	59
$34 \div 35$	22	24	$66 \div 67$	40	42	$98 \div 99$	57	60
$36 \div 37$	23	25	$68 \div 69$	41	44	$100 \div 101$	58	62
$38 \div 39$	24	26	$70 \div 71$	42	45			

Задача 334. *Проверить наличие автокорреляции для данных задачи 328 критерием Кенуя.*

Для $\tilde{y} = 37$ и упорядоченной последовательности $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ имеем последовательность пар точек, находящихся по одну сторону от медианы:

(1,51), (2,48); (26,41), (26,53); (33,51), (35,47); (44,31), (45,1);
(2,48), (5,51); (26,53), (30,58); (36,30), (40,19); (45,1), (46,21);
(5,51), (8,49); (30,58), (31,43); (40,19) (41,32); (46,21), (47,21);
(8,49), (10,50); (31,43), (33,43); (41,32) (42,11); (47,21), (48,19);
(13,33), (17,14); (33,43), (33,51); (42,11) (44,31); (48,19), (52,16).

Всего таких пар точек $N = 20$. В табл. 225 для $n = 28$ находим $n_{0,95} = 18$.

Так как $N = 20 > n_{0,95} = 18$, корреляция признается значимой.

5.2.2.1.8. Критерий Блума–Кифера–Розенблатта

Статистика критерия, предложенного Блумом, Кифером и Розенблаттом [570], строится следующим образом. Имеется совокупность точек (x_i, y_i), $i = 1, \ldots, n$. Через точку с координатами (x_i, y_i) проводятся прямые, параллельные осям координат (x, y), и подсчитывается количество точек $m_1(i)$, находящихся в первом квадранте (для которого $x_j > x_i$ и $y_j > y_i$), $m_2(i)$ — находящихся во втором квадранте (для которого $x_j < x_i$ и $y_j > y_i$), $m_3(i)$ — в третьем квадранте (для которого $x_j < x_i$ и $y_j < y_i$), $m_4(i)$ — в четвертом квадранте (для которого $x_j > x_i$ и $y_j < y_i$).

Статистикой критерия является величина

$$B = \frac{1}{n^3} \sum_{i=1}^{n} \left[m_1(i)\, m_4(i) - m_2(i)\, m_3(i) \right]^2 .$$

Критические значения $B(\alpha)$ при $n \to \infty$ $(n > 30)$ равны: $B(0{,}90) = 0{,}0469$; $B(0{,}95) = 0{,}0584$; $B\,(0{,}99) = 0{,}868$. При $B \geqslant B(\alpha)$ корреляция признается значимой с вероятностью α.

Задача 335. *Для совокупности $n = 10$ пар величин*

$$(x_i, y_i):\ (1,12),\ (2,17),\ (4,8),\ (12,14),\ (7,1),\ (2,4),\ (1,13),\ (13,6),\ (4,1),\ (10,9)$$

установить наличие корреляции критерием Блума–Кифера–Розенблатта при доверительной вероятности $\alpha = 0{,}95$.

Для пар (x_i, y_i) имеем следующие значения $m_1(i)$, $m_2(i)$, $m_3(i)$ и $m_4(i)$:

i	x_i	y_i	$m_1(i)$	$m_2(i)$	$m_3(i)$	$m_4(i)$	i	x_i	y_i	$m_1(i)$	$m_2(i)$	$m_3(i)$	$m_4(i)$
1	1	12	0	2	0	6	6	2	4	2	4	0	2
2	2	17	0	0	2	6	7	1	13	0	2	0	6
3	4	8	3	3	0	2	8	13	6	6	0	3	0
4	12	14	1	0	8	1	9	4	1	4	3	0	0
5	7	1	6	3	0	0	10	10	9	3	1	4	1

Статистика критерия равна

$$B = \frac{1}{10^3} \cdot \sum_{i=1}^{10} \left[m_1(i) \cdot m_4(i) - m_2(i) \cdot m_3(i)\right]^2 =$$
$$= 10^{-3} \cdot \left[(0 \cdot 6 - 2 \cdot 0)^2 + (0 \cdot 6 - 0 \cdot 2)^2 + \ldots + (3 \cdot 1 - 1 \cdot 4)^2\right] = 0{,}054.$$

Так как $B = 0{,}054 < B(0{,}95) = 0{,}0584$, с вероятностью $\alpha = 0{,}95$ гипотеза о наличии корреляции между x и y отклоняется.

5.2.2.2. Ранговая корреляция

Понимая под рангом выборочного значения случайной величины его номер в упорядоченной по возрастанию выборке, можно рассматривать для оценки силы связи случайных величин не их численные значения, а соответствующие им ранги.

Подробно методы ранговой корреляции изложены в работе Кендалла [422].

5.2.2.2.1. Коэффициент ранговой корреляции τ Кендалла

Предположим, имеется выборка пар случайных величин (x, y) объема n, которым соответствуют последовательности рангов R (для ряда x_i) и R^* (для ряда y_j). Расположим ряд значений x_i в порядке возрастания величины: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$. Тогда последовательность рангов R будет представлять собой последовательность натуральных чисел $1, 2, \ldots, n$. Значения y, соответствующие значениям x, образуют в этом случае некоторую последовательность рангов R^*.

Рассмотрим несколько способов оценки корреляции величин x и y, предложенных Кендаллом [422]. Назовем пару рангов R_j^* и R_ν^* $(j < \nu)$ инверсией, если в последовательности рангов R^* наблюдаем $R_j^* > R_\nu^*$ $(j = 1, \ldots, n-1)$. Обозначим через Q число таких пар. Тогда коэффициент корреляции, предложенный Кендаллом [422], равен

$$\tau = 1 - \frac{4Q}{n\,(n-1)}.$$

Второй способ вычисления τ заключается в следующем. В последовательности рангов R^* подсчитываем количество членов, расположенных справа от $R_j^* = 1$. Затем вычеркиваем $R_j^* = 1$ и подсчитываем число членов последовательности, расположенных справа от $R_j^* = 2$, и т. д. Обозначим сумму чисел, полученных с помощью

указанной процедуры, через K. Тогда τ вычисляется по формуле

$$\tau = \frac{4K}{n(n-1)} - 1.$$

Иногда используются эквивалентные формы записи τ:

$$\tau = \frac{2S}{n(n-1)}, \quad \text{где} \quad S = K - Q = 2K - \frac{n(n-1)}{2} = \frac{n(n-1)}{2} - 2Q.$$

Коэффициент τ принимает значения от -1 до $+1$. Равенство $\tau = 1$ указывает на строгую линейную корреляцию. При $n \geqslant 10$ распределения τ, S и K удовлетворительно аппроксимируются нормальным распределением с параметрами, соответственно:

$$\mathbf{M}(\tau) = 0; \quad \mathbf{D}(\tau) = \frac{2(2n+15)}{9n(n+1)}; \quad \mathbf{M}(S) = 0; \quad \mathbf{D}(S) = \frac{n(n-1)(2n+5)}{18};$$

$$\mathbf{M}(K) = \frac{n(n-1)}{4}; \quad \mathbf{D}(K) = \frac{n(n-1)(2n+5)}{72}.$$

Следовательно, при $n \geqslant 10$ наличие корреляции признается значимым с достоверностью α, если выполняется любое из следующих неравенств:

$$|\tau| > \tau_\alpha = u_\alpha \left\{ \frac{2\,(2n+5)}{9n\,(n-1)} \right\}^{\frac{1}{2}}; \quad |S| > S_\alpha = u_\alpha \left\{ \frac{n\,(n-1)\,(2n+5)}{18} \right\}^{\frac{1}{2}};$$

$$K \geqslant K_\alpha = \frac{n\,(n-1)}{4} + u_\alpha \left\{ \frac{n\,(n-1)\,(2n+5)}{72} \right\}^{\frac{1}{2}},$$

где u_γ — γ-квантиль стандартного нормального распределения.

Если среди значений x и y есть совпадающие значения (т. е. $x_i = x_\nu$ при $i \neq \nu$ или $y_j = y_\nu$ при $j \neq \nu$), то им приписываются средние ранги (например, если значения 3 и 4-го членов ранжированной выборки совпадают, то им приписывается одинаковый средний ранг $(3+4) : 2 = 3{,}5$). Если наблюдается q связей в ряду x и f связей в ряду y, то оценка τ корректируется следующим образом:

$$\tau = \frac{S}{\sqrt{\frac{1}{2}n(n-1) - T}\sqrt{\frac{1}{2}n(n-1) - U}},$$

где $T = \dfrac{1}{2}\sum\limits_{i=1}^{q} t_i\,(t_i - 1)$; $U = \dfrac{1}{2}\sum\limits_{i=1}^{f} u_i\,(u_i - 1)$; t_i (u_i) — длина i-й связи в ряду x (y).

В случае выборок из нормального распределения коэффициент τ может быть использован для быстрой оценки обычного коэффициента корреляции r по формуле $r = \sin\dfrac{\tau\pi}{2}$.

Задача 336. *Имеется последовательность пар* (x_i, y_i):

x_i:	2	4	7	1	5	9	11	12	17	8
y_i:	6	3	5	7	1	2	4	14	18	21.

Используя коэффициент корреляции Кендалла, установить наличие корреляционной зависимости между x *и* y *с достоверностью* $\alpha = 0{,}95$.

Упорядочим ряд значений x_i по возрастанию:

x_i:	1	2	4	5	7	8	9	11	12	17
y_j:	7	6	3	1	5	21	2	4	14	18.

Заменяя значения x_i и y_j их рангами, получаем последовательность рангов:

$$
\begin{array}{lcccccccccc}
R_i: & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
R_j^*: & 7 & 6 & 3 & 1 & 5 & 10 & 2 & 4 & 8 & 9.
\end{array}
$$

Далее находим для $R_1^* = 7$ число инверсий (когда $R_1^* > R_\nu^*$, $\nu > 1$) равно 6, для $R_2^* = 6 \to 5$, для $R_3^* = 3 \to 2$, для $R_4^* = 1 \to 0$, для $R_5^* = 5 \to 2$, для $R_6^* = 10 \to 4$, для $R_7^* = 2 \to 0$, для $R_8^* = 4 \to 0$, для $R_9^* = 8 \to 0$.

Таким образом, общее число инверсий равно $Q = 6 + 5 + 2 + 2 + 4 = 19$.

Следовательно, $\tau = 1 - \dfrac{4 \cdot 19}{10 \cdot 9} = 0{,}155$.

Теперь рассмотрим второй способ оценки τ. Для первоначальной последовательности рангов

$$R_j^*:\quad 7,\quad 6,\quad 3,\quad 1,\quad 5,\quad 10,\quad 2,\quad 4,\quad 8,\quad 9$$

определяем количество членов, находящихся справа от $R_4^* = 1$ — получаем 6 членов. Теперь вычеркиваем $R_4^* = 1$ и получаем ряд

$$R_j^*:\quad 7,\quad 6,\quad 3,\quad 5,\quad 10,\quad 2,\quad 4,\quad 8,\quad 9.$$

Справа от $R_6^* = 2$ находятся 3 члена. Вычеркиваем $R_6^* = 2$ и получаем ряд

$$R_j^*:\quad 7,\quad 6,\quad 3,\quad 5,\quad 10,\quad 4,\quad 8,\quad 9,$$

в котором справа от $R_3^* = 3$ находятся 5 членов. Далее, действуя по аналогии, находим

— в ряду R_j^*: 7, 6, 5, 10, 4, 8, 9 справа от $R_5^* = 4$ находятся 2 члена;
— в ряду R_j^*: 7, 6, 5, 10, 8, 9 справа от $R_3^* = 5$ находятся 3 члена;
— в ряду R_j^*: 7, 6, 10, 8, 9 справа от $R_2^* = 6$ находятся 3 члена;
— в ряду R_j^*: 7, 10, 8, 9 справа от $R_1^* = 7$ находятся 3 члена;
— в ряду R_j^*: 10, 8, 9 справа от $R_2^* = 8$ находится 1 член;
— в ряду R_j^*: 10, 9 справа от $R_2^* = 9$ находятся 0 членов.

Окончательно имеем $K = 6 + 3 + 5 + 2 + 3 + 3 + 3 + 1 + 0 = 26$.

Коэффициент τ равен $\tau = \dfrac{4 \cdot 26}{10 \cdot 9} - 1 = 0{,}155$, что и следовало ожидать.

Далее $S = K - Q = 26 - 19 = 7$ и $\tau = \dfrac{2 \cdot S}{n \cdot (n-1)} = \dfrac{2 \cdot 7}{10 \cdot 9} = 0{,}155$.

Для нормальной аппроксимации находим

$$\mathbf{D}(\tau) = \frac{2 \cdot (2n+5)}{9n \cdot (n-1)} = \frac{2 \cdot 25}{9 \cdot 10 \cdot 9} = 0{,}0617; \quad \mathbf{D}(S) = \frac{n \cdot (n-1) \cdot (2n+5)}{18} = \frac{10 \cdot 9 \cdot 25}{18} = 125;$$

$$\mathbf{D}(K) = \frac{n \cdot (n-1) \cdot (2n+5)}{72} = \frac{10 \cdot 9 \cdot 25}{72} = 31{,}25; \quad \mathbf{M}(K) = \frac{n \cdot (n-1)}{4} = \frac{10 \cdot 9}{4} = 22{,}5.$$

Далее для $u_{0,95} = 1{,}645$ имеем

$$\tau_{0,95} = 1{,}645 \cdot \sqrt{0{,}0617} = 0{,}409; \quad S_{0,95} = 1{,}645 \cdot \sqrt{125} = 18{,}39;$$

$$K_{0,95} = 22{,}5 + 1{,}645 \cdot \sqrt{31{,}25} = 31{,}69.$$

Так как $\tau = 0{,}155 < \tau_{0,95} = 0{,}409$; $S = 7 < S_{0,95} = 18{,}39$; $K = 26 < K_{0,95} = 31{,}69$, с вероятностью $\alpha = 0{,}95$ можно утверждать об отсутствии корреляции между x и y. Оценка обычного коэффициента корреляции равна

$$r = \sin\frac{\pi \cdot \tau}{2} = \sin\frac{\pi \cdot 0{,}155}{2} = 0{,}241.$$

5.2.2.2.2. Коэффициент корреляции ρ Спирмена

Рассматриваем последовательность рангов R_i (величин x_i) и R_j^* (величин y_j). Необходимости упорядочивать какую-либо совокупность рядов x_i и y_j нет. Находим разность рангов $d_i = R_i - R_j^*$, соответствующую паре (x_i, y_i). Коэффициент корре-

ляции Спирмена определяется формулой [422] $\rho = 1 - \dfrac{6\sum\limits_{i=1}^{n} d_i^2}{n(n^2-1)}$.

Его значения находятся в интервале от -1 до $+1$ ($\rho = 0$ указывает на отсутствие корреляции).

При $n \geqslant 10$ распределение ρ удовлетворительно описывается нормальным распределением с параметрами $\mathbf{M}(\rho) = 0$ и $\mathbf{D}(\rho) = \dfrac{1}{n-1}$.

Иногда в качестве статистики для проверки значимости ρ используется сумма квадратов отклонений рангов

$$S = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (R_i - R_i^*)^2 .$$

При $n \geqslant 10$ ее распределение также аппроксимируется нормальным распределением с параметрами

$$\mathbf{M}(S) = \frac{n(n^2-1)}{6}; \quad \mathbf{D}(S) = \frac{n^2(n+1)^2(n-1)}{36}.$$

Корреляция признается значимой при $|\rho| > \rho_\alpha$ или $S > S_\alpha$, где ρ_α и S_α — критические значения, равные при $n \geqslant 10$

$$\rho_\alpha = u_{\frac{1+\alpha}{2}} \frac{1}{\sqrt{n-1}}; \quad S_\alpha = \frac{n(n^2-1)}{6} + u_\alpha \frac{n(n+1)}{6}\sqrt{n-1}.$$

Более точная аппроксимация критических точек ρ предложена Иманом и Коновером [571]. В соответствии с их аппроксимацией используется статистика $J = \dfrac{\rho}{2}\left(\sqrt{n-1} + \sqrt{\dfrac{n-2}{1-\rho^2}}\right)$ критические значения которой равны $J(\alpha) = \dfrac{1}{2}u_\alpha + {}$ $+ \dfrac{1}{2}t_\alpha(n-2)$, где u_α — α-квантиль стандартного нормального распределения; t_α — α-квантиль распределения Стьюдента с $f = n-2$ степенями свободы. Если

$$J \geqslant J\left(\frac{1+\alpha}{2}\right) \quad \text{или} \quad J \leqslant -J\left(\frac{1+\alpha}{2}\right),$$

то гипотеза о наличии корреляции принимается с вероятностью α.

Для выборки из нормальных распределений может быть получена оценка для обычного коэффициента корреляции $r = 2\sin\dfrac{\pi}{6}\rho$.

В заключение приведем ряд полезных соотношений, связывающих между собой значения коэффициента корреляции τ и ρ: неравенство Дэниелса [422]

$$-1 \leqslant \frac{3(n+2)}{n-2}\tau - \frac{2(n+1)}{n-2}\rho \geqslant 1,$$

или при $n \to \infty$: $-1 \leqslant 3\tau - 2\rho \leqslant 1$;

неравенства Дарбина–Стюарта

при $r \geqslant 0$: $\quad \dfrac{3n\tau - (n-2)}{2(n+1)} \leqslant \rho \leqslant 1 - \dfrac{1-\tau}{2(n+1)}\left[(n-1)(1-\tau)+4\right]$

$$\left(\text{при } n \to \infty \quad \frac{3}{2}\tau - \frac{1}{2} \leqslant \rho \leqslant \frac{1}{2} + \tau - \frac{1}{2}\tau^2\right);$$

при $\tau < 0$: $\quad \dfrac{1}{2}\tau^2 + \tau - \dfrac{1}{2} \leqslant \rho \leqslant \dfrac{3}{2}\tau + \dfrac{1}{2}$.

Из приведенных соотношений следует, что хотя коэффициенты τ и ρ и связаны между собой, но эта связь не столь элементарна. На практике чаще всего, если значения обоих коэффициентов не слишком близки к единице, то $\rho \approx 1{,}5\tau$.

У читателя может возникнуть вопрос: стоит ли пользоваться коэффициентом τ, если вычисление коэффициента ρ значительно проще? Почему же тогда коэффициент τ применяется на практике чаще? Это связано с тем, что если необходимо учесть вновь поступившие значения случайных величин, то ρ в отличие от τ приходится рассчитывать заново по всем выборочным значениям.

Задача 337. *Используя данные задачи 336, проверить наличие корреляции с помощью коэффициента ρ Спирмена при доверительной вероятности $\alpha = 0{,}95$.*

Имеем последовательность рангов для x_i и y_i:

$$R_i:\ 2,\ 3,\ 5,\ 1,\ 4,\ 7,\ 8,\ 9,\ 10,\ 6;$$
$$R_i^*:\ 6,\ 3,\ 5,\ 7,\ 2,\ 2,\ 4,\ 8,\ 9,\ 10.$$

Вычисляем

$$\sum_{i=1}^{10}(R_i - R_i*)^2 = (2-6)^2 + (3-3)^2 + \ldots + (6-10)^2 = 120; \rho = 1 - \frac{6\cdot 120}{10\cdot 99} = 0{,}273.$$

При $\alpha = 0{,}95$ имеем $u_{\frac{1+\alpha}{2}} = u_{\frac{1+0{,}95}{2}} = u_{0{,}975} = 1{,}96$ и

$$\rho_{0{,}95} = \frac{1{,}96}{\sqrt{9}} = 0{,}653; \quad S_{0{,}95} = \frac{10\cdot 99}{6} + \frac{1{,}645\cdot 10\cdot 11\cdot\sqrt{9}}{6} = 255{,}47.$$

Так как $\rho = 0{,}273 < \rho_{0{,}95} = 0{,}653$ и $S = 120 < S_{0{,}95} = 255{,}47$, корреляция незначима. Рассмотрим теперь аппроксимацию Имана–Коновера. Находим

$$J = \frac{\rho}{2}\cdot\left(\sqrt{n-1} + \sqrt{\frac{n-2}{1-\rho^2}}\right) = \frac{0{,}273}{2}\cdot\left(\sqrt{9} + \sqrt{\frac{8}{1-0{,}273^2}}\right) = 0{,}811.$$

При $u_{0{,}975} = 1{,}96$ и $t_{0{,}975}(8) = 2{,}306$ имеем

$$J\left(\frac{1+\alpha}{2}\right) = J\,(0{,}975) = \frac{1{,}96}{2} + \frac{2{,}306}{2} = 2{,}133.$$

Так как $J = 0{,}811 < J(0{,}975) = 2{,}133$, гипотеза корреляции отклоняется.

Оценка обыкновенного коэффициента корреляции равна

$$r = 2\cdot\sin\frac{\pi\cdot 0{,}273}{6} = 0{,}285.$$

Легко убедиться теперь, что неравенства Дэниелса и Дарбина–Стюарта выполняются:

$$-1 \leqslant \frac{3\cdot(n+2)}{n-2}\cdot\tau - \frac{2\cdot(n+1)}{n-2}\cdot\rho \leqslant 1; \quad -1 \leqslant \frac{3\cdot 12\cdot 0{,}155}{8} - \frac{2\cdot 11\cdot 0{,}273}{8} = -0{,}0532 \leqslant 1;$$

$$\rho \leqslant 1 - \frac{1-\tau}{2\cdot(n+1)}\cdot[(n-1)(1-\tau)+4]; \quad 0{,}273 \leqslant 1 - \frac{1-0{,}155}{2\cdot 11}\cdot[9\cdot(1-0{,}155)+4] = 0{,}554;$$

$$\rho \geqslant \frac{3n\cdot\tau-(n-2)}{2\cdot(n+1)}; \quad 0{,}273 \geqslant \frac{3\cdot 10\cdot 0{,}155 - 8}{2\cdot 11} = -0{,}152.$$

5.2.2.2.3. Критерий Гёфдинга

Критерий Гёфдинга является ранговым аналогом критерия Блума–Кифера–Розенблатта (см. раздел 5.2.2.1.8). Статистика критерия строится следующим образом [572]: значения x_i и y_i предварительно ранжируются, а затем заменяются их рангами R_i и R_i^* соответственно.

Обозначим через C_i число пар из выборок (x_ν, y_ν), для которых одновременно $x_\nu < x_i$ и $y_\nu < y_i$:

$$C_i = \sum_{\substack{\nu=1\\ \nu\neq i}}^{n} \varphi(x_\nu, x_i)\,\varphi(y_\nu, y_i)\,, \quad i = 1, \ldots, n, \quad \text{где } \varphi(a,b) = \begin{cases} 1, & \text{при } a < b; \\ 1/2, & \text{при } a = b; \\ 0, & \text{при } a < b. \end{cases}$$

Положим

$$Q = \sum_{i=1}^{n} (R_i - 1)(R_i - 2)(R_i^* - 1)(R_i^* - 2)\,;$$

$$K = \sum_{i=1}^{n} C_i\,(R_i - 2)(R_i^* - 2)\,; \quad S = \sum_{i=1}^{n} C_i\,(C_i - 1)\,;$$

$$D = \frac{Q - 2(n-2)K + (n-2)(n-3)S}{n(n-1)(n-2)(n-3)(n-4)}.$$

Гипотеза корреляции принимается, если $D \geqslant D_\alpha$, где D_α — критическое значение, приведенное в табл. 226.

Таблица 226

Критические значения D_α критерия независимости Гёфдинга (α — доверительная вероятность) [18]

α	n				
	5	6	7	8	9
0,90	0,03330	–	0,00635	0,00476	0,00403
0,95	–	0,01660	0,00992	0,00773	0,00635
0,99	–	0,03330	0,01900	0,01488	0,01217

При $n \to \infty$ может быть использовано предельное распределение для критерия Блума–Кифера–Розенблатта B, исходя из того, что случайная величина $D + \dfrac{1}{36n}$ имеет такое же распределение, как и B.

Корреляция признается значимой:

— с вероятностью $\alpha = 0{,}90$, если $D > 0{,}0469 + \dfrac{1}{36n}$;

— с вероятностью $\alpha = 0{,}95$, если $D > 0{,}0584 + \dfrac{1}{36n}$;

— с вероятностью $\alpha = 0{,}99$, если $D > 0{,}0868 + \dfrac{1}{36n}$.

Задача 338. *Для выборочных пар данных*

$$(x_i, y_i):\quad (7,3),\quad (7,2),\quad (8,3),\quad (9,4),\quad (10,5),\quad (11,6)$$

установить наличие корреляции критерием Гёфдинга при доверительной вероятности $\alpha = 0{,}95$.

Находим

$$C_1 = \varphi(x_2,x_1)\cdot\varphi(y_2,y_1) + \varphi(x_3,x_1)\cdot\varphi(y_3,y_1) + \varphi(x_4,x_1)\cdot\varphi(y_4,y_1) + \\ + \varphi(x_5,x_1)\cdot\varphi(y_5,y_1) + \varphi(x_6,x_1)\cdot\varphi(y_6,y_1) = 0\cdot\frac{1}{2} + 0\cdot\frac{1}{2} + 0\cdot 0 + 0\cdot 0 + 0\cdot 0 = 0;$$

$$C_2 = \varphi(x_1,x_2)\cdot\varphi(y_1,y_2) + \varphi(x_3,x_2)\cdot\varphi(y_3,y_2) + \varphi(x_4,x_2)\cdot\varphi(y_4,y_2) + \\ + \varphi(x_5,x_2)\cdot\varphi(y_5,y_2) + \varphi(x_6,x_2)\cdot\varphi(y_6,y_2) = 0\cdot\frac{1}{2} + 0\cdot 0 + 0\cdot 0 + 0\cdot 0 + 0\cdot 0 = 0;$$

$$C_3 = \varphi(x_1, x_3) \cdot \varphi(y_1, y_3) + \varphi(x_2, x_3) \cdot \varphi(y_2, y_3) + \varphi(x_4, x_3) \cdot \varphi(y_4, y_3) + \\ + \varphi(x_5, x_3) \cdot \varphi(y_5, y_3) + \varphi(x_6, x_3) \cdot \varphi(y_6, y_3) = 1 \cdot \frac{1}{2} + 1 \cdot 1 + 0 \cdot 0 + 0 \cdot 0 + 0 \cdot 0 = 1{,}5.$$

Вычисляя дальше по аналогии, получаем $C_4 = 3$, $C_5 = 4$, $C_6 = 5$.

Теперь ранжируем ряд x_i:

x_i:	7	7	8	9	10	11
R_i:	1,5	1,5	3	4	5	6

и ряд y_i:

y_i:	3	2	3	4	5	6
R_i^*:	2,5	1	2,5	4	5	6.

(одинаковым значениям присвоены средние ранги).

Далее находим

$$Q = \sum_{i=1}^{6} (R_i - 1) \cdot (R_i - 2) \cdot (R_i^* - 1) \cdot (R_i^* - 2) =$$
$$= (1{,}5 - 1) \cdot (1{,}5 - 1) \cdot (2{,}5 - 1) \cdot (2{,}5 - 2) + \ldots + (6 - 1) \cdot (6 - 2) \cdot (6 - 1) \cdot (6 - 2) = 581{,}3125;$$

$$K = \sum_{i=1}^{6} C_i \cdot (R_i - 2) \cdot (R_i^* - 2) = 0 \cdot (1{,}5 - 2) \cdot (2{,}5 - 2) + \ldots + 5 \cdot (6 - 2) \cdot (6 - 2) = 128{,}75;$$

$$S = \sum_{i=1}^{6} C_i \cdot (c_i - 1) = 0 \cdot (-1) + 0 \cdot (-1) + 1{,}5 \cdot 0{,}5 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 4 = 38{,}75;$$

$$D = \frac{581{,}3125 - 2 \cdot 4 \cdot 128{,}25 + 4 \cdot 3 \cdot 38{,}75}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2} = 0{,}0282.$$

Из табл. 226 для $n = 6$ и $\alpha = 0{,}95$ находим $D_{0,95} = 0{,}0166$.

Так как $D = 0{,}0282 > D_{0,95} = 0{,}0166$, корреляция между x и y должна быть признана значимой.

5.2.2.2.4. Критерий Ширахатэ

Критерий Ширахатэ [573] является аналогом критерия Спирмена (см. раздел 5.2.2.2.2), асимптотически ему эквивалентен, но обладает большей эффективностью для малых выборок.

Определим ранги случайных величин (x_i, y_i), $i = 1, \ldots, n$, следующим образом:

$$R_i = \sum_{j=1}^{n} \{u(x_i - x_j) + u(x_i - y_j)\}; \quad R_{n+i} = \sum_{j=1}^{n} \{u(y_i - x_j) + u(y_i - y_j)\},$$

где $u(a) = \begin{cases} 1 & \text{при } x \geqslant 0, \\ 0 & \text{при } x < 0. \end{cases}$

Если случайные величины x_i и y_i коррелируют, то будут коррелировать и ранги R_i и R_{n+i}.

Статистикой критерия Ширахатэ является сумма

$$S = \sum_{i=1}^{n} R_i R_{n+i}.$$

Если бы мы определили ранги обычным способом:

$$R_i^* = \sum_{j=1}^{n} u(x_i - x_j) \quad \text{и} \quad R_{n+i}^* = \sum_{j=1}^{n} u(y_i - y_j),$$

то сумма $S^* = \sum\limits_{i=1}^{n} R_i^* R_{n+i}^*$ являлась бы статистикой Спирмена, уже рассмотренной ранее в разделе 5.2.2.2.2.

Статистики S и S^* асимптотически эквивалентны, но если значения статистики S^* находятся в интервале

$$\frac{n(n+1)(n+2)}{6} \leqslant S^* \leqslant \frac{n(n+1)(2n+1)}{6},$$

то статистка S ограничена интервалом

$$\frac{n(n+1)(2n+1)}{3} \leqslant S \leqslant \frac{n(n+1)(4n-1)}{3}.$$

Таким образом, если размах статистки S^* равен $\dfrac{n\left(n^2-1\right)}{6}$, то размах статистики S есть $\dfrac{2n(n^2-1)}{3}$, т. е. больше в 4 раза. Следовательно, S-критерий может дать больше информации, чем критерий, основанный на статистике S^*.

При $S_1(\alpha) < S < S_2(\alpha)$ корреляция признается незначимой (критические значения $S_1(\alpha)$ и $S_2(\alpha)$ приведены в табл. 227).

Таблица 227

Критические значения S_α критерия Ширахатэ [573]

n	Доверительная вероятность α				n	Доверительная вероятность α			
	0,90		0,95			0,90		0,95	
	S_1	S_2	S_1	S_2		S_1	S_2	S_1	S_2
4	62	96	61	97	10	911	1259	881	1289
5	116	175	114	181	12	1575	2125	1528	2173
6	199	195	194	302	14	2506	3314	2436	3386
7	313	456	302	468	16	3748	4877	3649	4977
8	467	666	450	683	18	5349	6863	5214	6999
9	664	932	642	955	20	7353	9311	7177	9501

Задача 339. *Используя данные и условия задачи* 336, *проверить наличие корреляции с помощью критерия Ширахатэ.*

Имеем данные (x_i, y_i):

$$\begin{array}{lllllllllll} x_i: & 1 & 2 & 4 & 5 & 7 & 8 & 9 & 11 & 12 & 17; \\ y_i: & 7 & 6 & 3 & 1 & 5 & 21 & 2 & 4 & 14 & 18. \end{array}$$

Для $i = 1$ находим

$$R_1 = \sum_{j=1}^{10} \{n\cdot(x_1 - x_j) + n\cdot(x_1 - y_j)\} = n\cdot(x_1 - x_1) + n\cdot(x_1 - y_1) + n\cdot(x_1 - x_2) + \\ + n\cdot(x_1 - y_2) + n\cdot(x_1 - x_3) + n\cdot(x_1 - y_3) + n\cdot(x_1 - x_4) + n\cdot(x_1 - y_4) + \ldots \\ \ldots + n\cdot(x_1 - x_10 + n\cdot(x_1 - y_10)) = \\ = 1+0+0+0+0+0+0+1+0+0+0+0+0+0+0+0+0+0+0+0 = 2.$$

Далее находим:

$$R_2 = 4;\quad R_3 = 7;\quad R_4 = 9;\quad R_5 = 12;\quad R_6 = 13;\quad R_7 = 14;\quad R_8 = 15;\quad R_9 = 16;\quad R_{10} = 18;$$

$$R_{10+1} = R_{11} + \sum_{j=1}^{10} [n\cdot(y_1 - x_j) + n\cdot(y_1 - y_j)] = \\ = 1+1+1+1+1+1+1+1+1+1+0+0+0+1+0+1+0+0+0+0 = 12;$$

$$R_{12} = 10;\quad R_{13} = 5;\quad R_{14} = 2;\quad R_5 = 9;\quad R_{16} = 20;\quad R_{17} = 4;\quad R_{18} = 7; \\ R_{19} = 17;\quad R_{20} = 19.$$

В результате получаем ряды:

R_i:	2	4	7	9	12	13	14	15	16	18;
R_{n+i}:	12	10	5	2	9	20	4	7	17	19.

Тогда

$$S = \sum_{i=1}^{n} R_i \cdot R_{i+1} = 2 \cdot 12 + 4 \cdot 10 + 7 \cdot 5 + 9 \cdot 2 + 12 \cdot 9 + 13 \cdot 20 + 14 \cdot 4 + 15 \cdot 7 + 16 \cdot 17 + 18 \cdot 19 = 1260.$$

Из табл. 227 для $n = 10$ и $\alpha = 0{,}95$ находим $S_1 = 881$ и $S_2 = 1289$.

Так как $S_1 = 881 < S = 1260 < S_2 = 1289$, корреляция признается незначимой.

5.2.2.2.5. Критерий корреляции Фишера–Йэйтса

Определяется формулой [574]

$$\rho* = \frac{\sum_{i=1}^{n} a_n(R_i)\, a_n(R_i*)}{\sum_{i=1}^{n} a_n^2(i)},$$

где $a_n(i)$ — математическое ожидание i-й порядковой статистики в выборке объема n из стандартного нормального распределения.

Значения ρ^* заключены между -1 и $+1$. Наибольшей эффективностью применение ρ^* обладает при исследовании зависимости между случайными величинами, имеющими распределение, близкое к нормальному.

При $n \geqslant 10$ распределение статистики

$$S^* = \sum_{i=1}^{n} a_n(R_i)\, a_n(R_i^*)$$

стремится к нормальному со средним $\mathbf{M}(S^*) = 0$ и дисперсией

$$\mathbf{D}(S^*) = \frac{1}{n-1} \left\{ \sum_{i=1}^{n} a_n^2(i) \right\}^2 .$$

Корреляция признается значимой с достоверностью α при

$$|S^*| > S(\alpha) u_{\frac{1+\alpha}{2}} \frac{1}{\sqrt{n-1}} \sum_{i=1}^{n} a_n^2(i).$$

Здесь, как и ранее, R_i и R_i^* обозначают ранги величин x_i и y_i в упорядоченных рядах x и y соответственно.

Для нахождения величин $a_n(i)$ можно пользоваться либо специальными таблицами (например, [24] — с. 150), либо аппроксимацией

$$a_n(i) \approx 4{,}91 \cdot \left\{ \left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}} \right)^{0,14} - \left(1 - \frac{i - \frac{3}{8}}{n + \frac{1}{4}} \right)^{0,14} \right\}.$$

Следует учитывать соотношение $a_n(i) = a_n(n + 1 - i)$, что значительно снижает необходимый объем вычислений.

Задача 340. *Используя данные задачи 336, проверить наличие корреляции с помощью критерия Фишера–Йэйтса.*

Для наших данных

x_i:	1	2	4	5	7	8	9	11	12	17
y_i:	7	6	3	1	5	21	2	4	14	18

последовательность рангов будет иметь вид:

R_i:	1	2	3	4	5	6	7	8	9	10
R_i^*:	7	6	3	1	5	10	2	4	8	9

Используя аппроксимацию

$$a_{10}(i) = 4{,}91 \cdot \left[\left(\frac{i - \frac{3}{8}}{10 + \frac{1}{4}}\right)^{0,14} - \left(1 - \frac{i - \frac{3}{8}}{10 + \frac{1}{4}}\right)^{0,14}\right],$$

вычисляем значения $a_{10}(R_i)$ и $a_{10}(R_i^*)$. Результаты сводим в таблицу:

i	R_i	R_i^*	$a_{10}(R_i)$	$a_{10}(R_i^*)$	i	R_i	R_i^*	$a_{10}(R_i)$	$a_{10}(R_i^*)$
1	1	7	−1,547980	0,373926	6	6	10	0,122033	1,547980
2	2	6	−0,998750	0,122033	7	7	2	0,373926	−0,998750
3	3	3	−0,653292	−0,653292	8	8	4	0,653292	0,373926
4	4	1	−0,373926	−1,547980	9	9	8	0,998750	0,653292
5	5	5	−0,122033	−0,122033	10	10	9	1,547980	0,998750

Вычисляем

$$S^* = \sum_{i=1}^{10} a_{10}(R_i) \cdot a_{10}(R_i^*) = (-1{,}54798) \cdot 0{,}373926 + \ldots + 1{,}54798 \cdot 0{,}99875 = 2{,}44965;$$

$$\sum_{i=1}^{10} a_{10}^2(i) = 7{,}95049; \quad \rho^* = \frac{2{,}44965}{7{,}95049} = 0{,}308.$$

Для $\alpha = 0{,}95$ имеем $u_{0,975} = 1{,}96$ и $S(0{,}95) = \dfrac{u_{0,975} \cdot \sum_{i=1}^{10} a_{10}^2(i)}{\sqrt{n-1}} = 5{,}194.$

Так как $S^* = 2{,}44965 < S(0{,}95) = 5{,}194$, корреляция признается незначимой.

5.2.2.2.6. Коэффициент корреляции Ван дер Вардена

Определяется формулой [365] $\tilde{\rho} = \dfrac{\sum_{i=1}^{n} u_{\frac{R_i}{n+1}} u_{\frac{R_i^*}{n+1}}}{\sum_{i=1}^{n} u^2_{\frac{i}{n+1}}}$, где u_α — α-квантиль стандартного нормального распределения. Коэффициент оптимален для оценки корреляции нормально распределенных случайных величин.

При $n \geqslant 10$ распределение $\tilde{S} = \sum_{i=1}^{n} u_{\frac{R_i}{n+1}} u_{\frac{R_i^*}{n+1}}$ стремится к нормальному с параметрами

$$\mathbf{M}(\tilde{S}) = 0; \quad \mathbf{D}(\tilde{S}) = \frac{1}{n-1}\left\{\sum_{i=1}^{n} u^2_{\frac{i}{n+1}}\right\}^2.$$

Отсюда: если $|\tilde{S}| > \tilde{S}(\alpha) = u_{\frac{1+\alpha}{2}} \dfrac{1}{\sqrt{n-1}} \sum\limits_{i=1}^{n} u^2_{\frac{i}{n+1}}$, то с вероятностью α корреляция признается значимой. Для поиска u_α могут быть использованы таблицы, либо аппроксимации (см. раздел 1.1.1), например,

$$u_\alpha \approx 4{,}91\left[\alpha^{0,14} - (1-\alpha)^{0,14}\right].$$

Следует помнить, что $u_\alpha = -u_{1-\alpha}$.

Задача 341. *Используя данные задачи 336, проверить наличие корреляции с помощью критерия Ван дер Вардена.*

Используя аппроксимацию $u_{\frac{i}{n+1}} = 4{,}91 \cdot \left\{\left(\dfrac{i}{n+1}\right)^{0,14} - \left(1 - \dfrac{i}{n+1}\right)^{0,14}\right\}$, вычисляем значения $u_{\frac{R_i}{11}}$ и $u_{\frac{R_i*}{11}}$ (значения R_i и R_i^* используем из задачи 340). Результаты сводим в таблицу:

i	R_i	R_i^*	$u_{\frac{R_i}{11}}$	$u_{\frac{R_i^*}{11}}$	i	R_i	R_i^*	$u_{\frac{R_i}{11}}$	$u_{\frac{R_i^*}{11}}$
1	1	7	−1,335080	0,347309	6	6	10	0,113674	1,335080
2	2	6	−0,906470	0,113674	7	7	2	0,347309	−0,906470
3	3	3	−0,602509	−0,602509	8	8	4	0,602509	−0,347309
4	4	1	−0,347309	−1,335080	9	9	8	0,906470	0,602509
5	5	5	−0,113674	−0,113674	10	10	9	1,335080	0,906470

Вычисляем

$$\tilde{S} = \sum_{i=1}^{10} u_{\frac{R_i}{n+1}} \cdot u_{\frac{R_i*}{n+1}} = (-1{,}33508)\cdot 0{,}347309 + (-0{,}90647)\cdot 0{,}113674 + \ldots$$
$$\ldots + 1{,}33508 \cdot 0{,}90647 = 1{,}65687;$$

$$\sum_{i=1}^{10} u^2_{\frac{i}{n+1}} = 6{,}201377; \quad \tilde{\rho} = \frac{1{,}65687}{6{,}201373} = 0{,}26718; \quad \tilde{S}\,(0{,}95) = \frac{u_{0,975}\cdot 6{,}201377}{3} = 4{,}051.$$

Так как $\tilde{S} = 1{,}656 < \tilde{S}\,(0{,}95) = 4{,}051$, гипотеза о наличии корреляции отклоняется с вероятностью 0,95.

5.2.2.2.7. Коэффициент конкордации Кендалла–Бэбингтона Смита

До сих пор рассматривалась корреляция двух случайных величин. Часто возникает необходимость исследовать корреляцию нескольких последовательностей значений случайных величин. Предположим, имеется k последовательностей рангов с равным числом рангов n в каждой последовательности

$$\begin{array}{cccccc}
R_{11} & R_{12} & \ldots & R_{1i} & \ldots & R_{1n}; \\
R_{21} & R_{22} & \ldots & R_{2i} & \ldots & R_{2n}; \\
. & . & \ldots & . & \ldots & . \\
. & . & \ldots & . & \ldots & . \\
. & . & \ldots & . & \ldots & . \\
R_{j1} & R_{j2} & \ldots & R_{ji} & \ldots & R_{jn}; \\
. & . & \ldots & . & \ldots & . \\
. & . & \ldots & . & \ldots & . \\
. & . & \ldots & . & \ldots & . \\
R_{k1} & R_{k2} & \ldots & R_{ki} & \ldots & R_{kn}.
\end{array}$$

В качестве меры связи k последовательностей Кендалл и Б. Смит [575] предложили коэффициент конкордации (согласованности)

$$W = \frac{12S_W}{k^2\left(n^3-n\right)}, \quad \text{где} \quad S_W = \sum_{i=1}^{n}\left\{\sum_{j=1}^{k} R_{ji} - \frac{k\,(n+1)}{2}\right\}^2.$$

Легко видеть, что S_W является суммой отклонений рангов от их среднего значения. Значения W располагаются в интервале от 0 до 1.

Для $n \geqslant 20$ величины W и S_W распределены приблизительно нормально [25] с параметрами

$$\mathbf{M}(W) = \frac{1}{k}; \quad \mathbf{D}(W) = \frac{2(k-1)}{k^3(n-1)};$$

$$\mathbf{M}(S_W) = \frac{k(n^3-n)}{12}; \quad \mathbf{D}(S_W) = \frac{k(k-1)(n+1)(n^2-1)n^2}{72}.$$

В силу несимметричности распределение W при $n \geqslant 20$ лучше аппроксимируется бета-распределением [25].

Точные критические суммы $S_W(\alpha)$ приведены в табл. 228. Если $S_W > S_W(\alpha)$, то наличие согласованности признается значимым с вероятностью α.

Таблица 228

Критические значения $S_W(\alpha)$ для коэффициента конкордации W [422]

k	Доверительная вероятность $\alpha = 0{,}95$					Доверительная вероятность $\alpha = 0{,}99$				
	n					n				
	3	4	5	6	7	3	4	5	6	7
3			64,4	103,9	157,3			75,6	122,8	185,6
4		49,5	88,4	143,3	217,0		61,4	109,3	176,2	265,0
5		62,6	112,3	182,4	276,2		80,5	142,8	229,4	343,8
6		75,7	136,1	281,4	335,2		99,5	176,1	282,4	422,6
8	48,1	101,7	183,7	299,0	453,1	66,8	137,4	242,7	388,3	579,9
10	60,0	127,8	231,2	376,7	571,0	85,1	175,3	309,1	494,0	737,0
15	89,8	192,9	349,8	570,5	864,9	131,0	269,8	475,2	758,2	1129,5
20	119,7	258,0	468,5	764,4	1158,7	177,0	364,2	641,2	1022,2	1521,9

При $n > 10 \div 15$ и отсутствии корреляции величина $k(n-1)W$ распределена приблизительно как χ^2 с $f = n-1$ степенями свободы. Отсюда следует, что критическое значение равно $W_\alpha = \dfrac{\chi^2_\alpha}{k(n-1)}$.

Если $W > W_\alpha$, то с вероятностью α корреляция между изучаемыми последовательностями признается значимой.

Если среди последовательностей рангов есть совпадения, то коэффициент конкордации следует вычислять по формуле

$$W = \frac{12S_W}{k^2\left(n^2-1\right) - k\sum\limits_{j=1}^{k} T_j}, \quad \text{где}$$

$T_j = \sum\limits_{t_j}\left(t_j^3 - t_j\right)$, t_j — количество совпавших рангов в j-й последовательности.

Совпавшим рангам, как и ранее, присваиваются средние ранги.

Задача 342. *Предположим, что имеются $k = 4$ последовательности числовых рядов, объемом $n = 10$ каждая:*

x_i:	1	3	7	9	12	14	18	19	21	26;
y_i:	7	8	6	1	4	2	7	0	3	1;
z_i:	11	12	7	6	5	4	3	1	0	−1;
l_i:	−1	0	1	12	4	5	7	3	2	−4.

Необходимо проверить согласованность рядов с помощью коэффициента конкордации Кендалла–Б. Смита при доверительной вероятности $\alpha = 0{,}95$.

Имеем совокупность рангов R_{ji} и квадратов вида

$$\oplus = \left(\sum_{j=1}^{k} R_{ij} - \frac{k\cdot(n+1)}{2}\right)^2 = \left(\sum_{j=1}^{4} R_{ij} - 22\right)^2$$

i	R_{ij}				$\oplus$
	j				
	1	2	3	4	
1	1	8,5	9	2	2,25
2	2	10	10	3	9
3	3	7	8	4	0
4	4	2,5	7	10	2,25
5	5	6	6	7	4
6	6	4	5	8	1
7	7	8,5	4	9	42,25
8	8	1	3	6	16
9	9	5	2	5	1
10	10	2,5	1	1	56,25

Далее $S_W = \sum_{i=1}^{10}\left\{\sum_{j=1}^{4} R_{ji} - \frac{k\cdot(n+1)}{2}\right\}^2 = 134$; $W = \frac{12\cdot 134}{4^2\cdot(1000-10)} = 0{,}101$.

Имея в виду, что $\chi^2_{0,95}(9) = 16{,}919$ (см. табл. 55), получаем

$$W_{0,95} = \frac{\chi^2_{0,95}}{4\cdot 9} = \frac{16{,}919}{36} = 0{,}470.$$

Так как $W = 0{,}101 < W_{0,95} = 0{,}470$, с вероятностью $\alpha = 0{,}95$ можно признать согласованность рангов незначимой.

5.2.2.2.8. Коэффициент конкордации Шукени–Фроли

Для случая двух групп экспертов Шукени и Фроли [576] предложили аналог коэффициента конкордации Кендалла–Б. Смита (см. раздел 5.2.2.2.7). Пусть две группы экспертов численностями m и n ставят перед собой задачу проранжировать k объектов. Обозначим через R_{ij} ($i = 1, \ldots, m$; $j = 1, \ldots, k$) — ранги, предложенные m экспертами первой группы; через R^*_{ij} ($1 = 1, \ldots, n$; $j = 1, \ldots, k$) — ранги, предложенные n экспертами второй группы ($R_j = \sum R_{ij}$ и $R^*_j = \sum R^*_{ij}$).

Статистика Шукени–Фроли равна

$$L = \sum_{j=1}^{k} R_j R^*_j.$$

Значение статистики L находится в интервале [455]

$$\frac{mnk\,(k+1)\,(k+2)}{6} \leqslant L \leqslant \frac{mnk\,(k+1)\,(2k+1)}{6}.$$

Известно [455], что

$$\mathbf{M}(L) = \frac{mnk(k+1)^2}{4}; \quad \mathbf{D}(L) = \frac{mn(k-1)k^2(k+1)^2}{144}.$$

Обобщенный коэффициент конкордации Шукени–Фроли определяется соотношением $\tilde{W} = \dfrac{L - \mathbf{M}(L)}{L_{\max} - \mathbf{M}(L)}$.

Если ρ_{ij} — коэффициент корреляции Спирмена (см. раздел 5.2.2.2.2) для i-го эксперта первой группы и j-го эксперта второй группы, то [455] $\tilde{W} = \dfrac{1}{mn}\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m}\rho_{ij}$.

Предельное распределение коэффициента Шукени–Фроли отлично от нормального и неудобно для применения. Значение $\tilde{W}$ вблизи $+1$ означает высокую степень согласованности внутри обеих групп экспертов и между группами; вблизи -1 — высокую степень согласия внутри групп и сильное несогласие между группами; вблизи 0 — либо несогласие внутри групп экспертов, либо согласие внутри групп экспертов при несогласии между ними.

Задача 343. *Две группы экспертов в количествах $m = 6$ и $n = 8$ провели ранжирование $k = 5$ объектов (результаты приведены в таблице). Необходимо проверить согласованность мнений экспертов критерием Шукени–Фроли.*

Номер эксперта	Ранжировка объектов				
Группа 1					
1	1	3	4	2	5
2	1	2	3	4	5
3	4	3	2	1	5
4	1	2	3	4	5
5	2	1	3	4	5
6	5	4	3	2	1
$\sum$	$R_1 = 14$	$R_2 = 15$	$R_3 = 18$	$R_4 = 17$	$R_5 = 26$
Группа 2					
1	1	2	3	4	5
2	3	2	1	5	4
3	4	5	1	2	3
4	1	2	3	4	5
5	5	4	2	3	1
6	1	2	3	4	5
7	3	2	4	5	1
8	1	5	4	3	2
$\sum$	$R_1^* = 19$	$R_2^* = 24$	$R_3^* = 21$	$R_4^* = 30$	$R_5^* = 26$

Вычисляем $L = \sum\limits_{j=1}^{5} R_j \cdot R_j^* = 14 \cdot 19 + 15 \cdot 24 + 18 \cdot 21 + 17 \cdot 30 + 26 \cdot 26 = 2190$.

Границы изменения L: $\dfrac{6 \cdot 8 \cdot 5 \cdot 6 \cdot 7}{6} = 1680 \leqslant L \leqslant 5040 = \dfrac{6 \cdot 8 \cdot 5 \cdot 6 \cdot 21}{6}$.

Далее вычисляем

$$\mathbf{M}(L) = \frac{6\cdot 8\cdot 5\cdot 6^2}{4} = 2160; \quad \mathbf{D}(L) = \frac{6\cdot 8\cdot 4\cdot 25\cdot 36}{144} = 1200 \quad (\sqrt{\mathbf{D}(L)} = 34{,}64);$$

$$\tilde{W} = \frac{2190-2160}{5040-2160} = 0{,}00729.$$

Так как $\tilde{W} = 0{,}00729 \approx 0$, согласованность внутри группы экспертов либо между ними очень низка.

5.2.2.3. Точечно-бисериальная корреляция

При проведении некоторых исследований часто сталкиваются с проблемой выяснения взаимосвязи между характеристиками, одна из которых может быть ранжирована, а вторая допускает только группировку в две группы по качественному признаку (дихотомия).

Приведем известные результаты для оценки зависимости в такой ситуации. Одной из таких оценок является коэффициент точечно-бисериальной корреляции

$$r_\delta = \frac{\bar{x}_1 - \bar{x}_2}{S}\sqrt{\frac{n_1 n_2}{n(n+1)}},$$

где n_1, n_2 — объемы двух групп; $n = n_1 + n_2$; $\bar{x}_1$, $\bar{x}_2$ — средние значения первой характеристики в двух группах, образованных в соответствии со второй характеристикой;

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n}; \quad S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2; \quad S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(x_i - \bar{x}_2)^2.$$

Значимость корреляции определяется так же, как и для обычного коэффициента корреляции (см. раздел 5.2.1.1). Если $|r_\delta| > r_\alpha$, то корреляция признается значимой.

Рассмотрим теперь аналогичную задачу для ранговой корреляции, применительно к коэффициенту ранговой корреляции τ Кендалла (см. раздел 5.2.2.2.1). Напомним, что в случае отсутствия дихотомии

$$\tau = \frac{2S}{n(n-1)}.$$

Предположим, что дихотомия реализуется в разделении данных по двум признакам в группы объемов n_1 и n_2 $(n_1 + n_2 = n)$.

В этом случае коэффициент точечно-бисериальной корреляции Кендалла равен

$$r_\delta = S\left\{\frac{1}{4}n(n-1)[n(n-1) - n_1(n_1-1) - n_2(n_2-1)]\right\}^{\frac{1}{2}}.$$

Проверка значимости τ_δ проводится аналогично проверке значимости τ (см. раздел 5.2.2.2.1).

Задача 344. *В нашем распоряжении имеются следующие данные (x_i — количественный признак, y — качественный признак, обозначаемый символом + или −):*

x_i:	1	2	3	4	5	6	7	8	9	10;
y:	+	+	−	−	−	+	−	+	−	−.

Необходимо оценить наличие зависимости между x и y при достоверности $\alpha = 0{,}95$.

Используем коэффициент точечно-бисериальной корреляции r_δ. Имеем $n_1 = 4$, $n_2 = 6$ $(n_1 + n_2 = n = 10)$. Находим далее

$$\bar{x}_1 = \frac{1+2+6+8}{4} = 4{,}25; \quad \bar{x}_2 = \frac{3+4+5+7+9+10}{6} = 6{,}33;$$

$$S_1^2 = \frac{1}{3}\sum_{i=1}^{4}(x_i - \bar{x}_1)^2 = 10{,}916; \quad S_2^2 = \frac{1}{5}\sum_{i=1}^{6}(x_i - \bar{x}_2)^2 = 9{,}085; \quad S = 3{,}014;$$

$$r_\delta = \frac{4{,}25 - 6{,}33}{3{,}014}\sqrt{\frac{4 \cdot 6}{110}} = -0{,}322.$$

Для $\alpha = 0{,}95$ и $n = 10$ из табл. 217 имеем $r_{0,95} = 0{,}632$.

Так как $|r_\delta| = 0{,}322 < r_{0,95} = 0{,}632$, корреляция признается незначимой.

Теперь вычислим точечно-бисериальный коэффициент корреляции Кендалла (см. раздел 5.2.2.2.1). Имеем последовательность рангов

$$\begin{array}{lcccccccccc} R_j\colon & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ R_j^*\colon & 2,5 & 2,5 & 7,5 & 7,5 & 7,5 & 2,5 & 7,5 & 2,5 & 7,5 & 7,5 \end{array}$$

Поясним подсчет рангов R_j^*. Имеем последовательность членов

$$+ \; + \; - \; - \; - \; + \; - \; + \; - \; -,$$

что эквивалентно наличию $n_1 = 4$ и $n_2 = 6$ равных элементов, которым приписываем равные средние ранги. Элементы $+$, имеющие (условно) номера 1, 2, 3 и 4, получают равный средний ранг $\dfrac{1+2+3+4}{4} = 2{,}5$. Аналогично и элементы $-$ получают равные ранги $\dfrac{5+6+7+8+9+10}{6} = 7{,}5$. Теперь подсчитаем сумму Q по алгоритму, изложенному в разделе 5.2.2.2.1 (т. е. число инверсий, когда $R_j^* < R_\nu^*$): $Q = 6 + 6 + 3 + 2 = 17$.

Далее вычисляем

$$S = \frac{n \cdot (n-1)}{2} - 34 = \frac{10 \cdot 9}{2} - 34 = 11; \quad \tau_\delta = \frac{11}{\sqrt{\frac{1}{4} \cdot 10 \cdot 9 \cdot (10 \cdot 9 - 4 \cdot 3 - 6 \cdot 5)}} = 0{,}335;$$

$$\tau_{0,95} = u_{0,95} \cdot \sqrt{\frac{2 \cdot (2n+5)}{9n \cdot (n-1)}} = 1{,}645 \cdot \sqrt{\frac{2 \cdot 25}{90 \cdot 9}} = 0{,}409.$$

Так как $\tau_\delta = 0{,}335 < \tau_{0,95} = 0{,}409$, точечно-бисериальная корреляция незначима.

5.2.2.4. Статистическая оценка связи между качественными признаками (таблицы сопряженности признаков)

Предположим, что наблюдаемая случайная величина может изменяться в зависимости от некоторых признаков — например, долговечность электронного прибора может зависеть от технологии изготовления, применяемых материалов. По результатам наблюдений над случайной величиной, классифицированным по наличию или отсутствию исследуемых признаков, необходимо ответить на вопрос, существует ли взаимосвязь между ними, иными словами: связано ли обладание одним признаком с обладанием другим признаком (в условиях приведенного примера — существует ли связь между технологией изготовления и применяемыми материалами).

По существу, в данном случае мы имеем переход от точечно-бисериальной корреляции к изучению зависимости между двумя (или несколькими) качественными признаками.

Таблицы, в которых представлены значения исследуемой случайной величины, классифицированные по качественным признакам, называются таблицами сопряженности признаков.

5.2.2.4.1. Оценка связи признаков в таблицах сопряженности 2 × 2

Если исследуется взаимосвязь двух признаков A и B, то таблица сопряженности называется таблицей 2×2, или четырехклеточной таблицей и имеет вид

	a	b	
	c	d	,

где a — число элементов выборки, обладающим признаками A и B одновременно; b — число элементов выборки, обладающих признаком A, но не обладающих признаком B; c — число элементов выборки, обладающих признаком B, но не обладающих признаком A; d — число элементов выборки, не обладающих ни одним из признаков A и B.

5.2.2.4.1.1. Меры связи в таблицах сопряженности 2 × 2

Рассмотрим три известные меры связи, позволяющие грубо оценить ее наличие.

5.2.2.4.1.1.1. Коэффициент ассоциации [132]

Находится по формуле

$$Q = \frac{ad - bc}{ad + bc}.$$

Если признаки A и B независимы, то $Q = 0$. В случае полной связи между признаками $Q = \pm 1$. Дисперсия Q равна [132]

$$\mathbf{D}(Q) = \frac{1}{4}\left(1 - Q^2\right)\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right).$$

Сравнение Q с полученным значением дисперсии (с учетом масштаба $\sqrt{\mathbf{D}(Q)}$) позволяет получить хотя бы первое приближение по оценке связи.

5.2.2.4.1.1.2. Коэффициент коллигации Юла [132]

Находится по формуле

$$K = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \quad \text{с дисперсией} \quad \mathbf{D}(K) = \frac{1}{16}\left(1 - K^2\right)\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right).$$

Между Q и K существует связь

$$Q = \frac{2K}{1 + K^2}.$$

Задача 345. *Предположим, что после перепроверки партии электронных ламп из $n = 110$ шт., изготовленных по двум технологиям, получены результаты, приведенные в таблице. Необходимо оценить связь качества ламп с технологией их изготовления с помощью оценок, изложенных в разделе* 5.2.2.4.1.1.

Технология	Число дефектных ламп	Число годных ламп	Всего ламп
1	95	15	110
2	70	40	110
Итого	165	55	220

В нашем случае имеем $a = 95$, $b = 15$, $c = 70$, $d = 40$.
Вычисляем коэффициент ассоциации

$$Q = \frac{95 \cdot 40 - 15 \cdot 70}{95 \cdot 40 + 15 \cdot 70} = 0{,}567;$$

$$\mathbf{D}(Q) = \frac{1}{4} \cdot \left(1 - 0{,}567^2\right) \cdot \left(\frac{1}{95} + \frac{1}{15} + \frac{1}{70} + \frac{1}{40}\right) = 0{,}019758 \quad \left(\sqrt{\mathbf{D}(Q)} = 0{,}140\right).$$

Находим коэффициент коллигации Юла

$$K = \frac{\sqrt{95 \cdot 40} - \sqrt{15 \cdot 70}}{\sqrt{95 \cdot 40} + \sqrt{15 \cdot 70}} = 0{,}311;$$

$$\mathbf{D}(K) = \frac{1}{16} \cdot \left(1 - 0{,}311^2\right) \cdot \left(\frac{1}{95} + \frac{1}{15} + \frac{1}{70} + \frac{1}{40}\right) = 0{,}00675 \quad (\sqrt{\mathbf{D}(K)} = 0{,}082).$$

Из анализа полученных оценок можно сделать оценочный вывод о том, что по вероятности существует связь между технологией изготовления и качеством продукции, так как коэффициенты $Q = 0{,}567$ и $K = 0{,}311$ достаточно велики по сравнению со своими среднеквадратическими отклонениями (превышают их более, чем в три раза).

5.2.2.4.1.1.3. Коэффициент контингенции (сходства) [132]

Основан на формуле

$$V = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}.$$

На практике для проверки гипотезы о существовании взаимосвязи между изучаемыми признаками используется величина $\chi^2 = nV^2$, имеющая при отсутствии связи распределение χ^2 с $f = 1$ степенью свободы.

С учетом поправки на непрерывность статистика критерия контингенции для проверки связи признаков имеет вид

$$\chi^2 = \frac{n\left(|ad - bc| - \dfrac{n}{2}\right)^2}{(a+b)\,(a+c)\,(b+d)\,(c+d)},$$

где n — общее число данных в таблице: $n = a + b + c + d$.

Если $\chi^2 > \chi^2_\alpha(1)$, то с достоверностью α зависимость признаков A и B признается значимой. Приведем некоторые полезные значения: $\chi^2_{0.90}(1) = 2{,}70$; $\chi^2_{0.95}(1) = 3{,}85$; $\chi^2_{0.99}(1) = 6{,}58$.

Задача 346. *Проверить значимость связи признаков в таблице сопряженности* 2×2 *в условиях задачи* 344 *коэффициентом контингенции при доверительной вероятности* $\alpha = 0{,}95$.

Имеем $V = \dfrac{95 \cdot 40 - 15 \cdot 70}{\sqrt{110 \cdot 165 \cdot 55110}} = 0{,}262$.

Далее при $n = 220$ имеем $\chi^2 = \dfrac{220 \cdot (95 \cdot 40 - 15 \cdot 70 - 110)^2}{100 \cdot 165 \cdot 55 \cdot 110} = 14{,}07$.

Так как $\chi^2 = 14{,}07 > \chi^2_{0{,}95}(1) = 3{,}85$, связь признаков в задаче должна быть признана значимой.

5.2.2.4.1.1.4. Точный критерий Фишера

Критерий $\chi^2_\alpha = nV^2$ применим при $n \geqslant 40$ и $a, b, c, d \geqslant 5$. Если эти условия не выполняются, то следует воспользоваться точным критерием Фишера, основанном на статистике

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{(a+b+c+d)!} \sum_{i=0}^{a} \frac{1}{(a+b-i)!(a+c-i)!(a+d-i)!}$$

Если $p > 1 - \alpha$, то с достоверностью α связь признаков признается значимой.

Задача 347. *Имеется следующая таблица сопряженности признаков 2×2:*

			$\sum$
	11	6	17
	13	10	23
$\sum$	24	16	40

Необходимо при доверительной вероятности $\alpha = 0{,}95$ оценить значимость связи признаков в таблице точным критерием Фишера.

Вычисляем

$$p = \frac{(11+6)!\,(13+10)!\,(11+13)!\,(6+10)!}{(11+6+13+10)!} \sum_{i=0}^{11} \frac{1}{i!\,(11+6-i)!\,(11+13-i)!\,(11+10-i)!} = 0{,}227.$$

Так как $p = 0{,}227 > 1-\alpha = 1-0{,}95 = 0{,}05$, связь признаков в таблице следует считать значимой.

5.2.2.4.1.1.5. Быстрые критерии оценки связи в таблицах сопряженности 2 × 2

При $a+b+c+d \geqslant 25$ и при условии $a+b = c+d$ или $a+c = b+d$ критерием является величина

$$z = \frac{(a+d)-(b+c)}{\sqrt{a+b+c+d}}.$$

Связь признаков в таблице с достоверностью α признается значимой при $z > u_\alpha$ (u_α — α-квантиль стандартного нормального распределения).

При $a+b \geqslant 10$ при условии $a+b \ll c+d$ или $a+c \ll b+d$ справедлив критерий

$$\tilde z = \frac{a-b+\dfrac{(a+c-b-d)(a+b)}{a+b+c+d}}{\sqrt{a+b}}.$$

Если $\tilde z > u_\alpha$, то связь признаков признается значимой.

Задача 348. *Для таблицы сопряженности*

			$\sum$
	14	28	42
	6	36	42
$\sum$	20	64	84

проверить гипотезу о согласованности признаков при доверительной вероятности $\alpha = 0{,}95$.

Имеем $a = 14$, $b = 28$, $c = 6$, $d = 36$. Так как $a+b = 42 = c+d$, используем критерий

$$z = \frac{(14+36)-(28+6)}{\sqrt{84}} = 1{,}745.$$

Так как $z = 1{,}745 > u_{0{,}95} = 1{,}645$, связь признаков следует признать значимой. Следует помнить, что всегда при z берется знак + (это не меняет результат).

Задача 349. *Для таблицы сопряженности*

			$\sum$
	15	4	19
	85	77	162
$\sum$	100	81	181

проверить гипотезу о согласованности признаков при доверительной вероятности $\alpha = 0{,}95$.

В нашем случае $a + b = 19 \ll c + d = 162$, поэтому используем критерий

$$\tilde{z} = \frac{15 - 4 - \dfrac{(15+85-77)\cdot(15+4)}{15+4+85+77}}{\sqrt{15+4}} = 1{,}97.$$

Так как $\tilde{z} = 1{,}97 > u_{0,95} = 1{,}645$, связь признаков в таблице следует признать значимой.

5.2.2.4.1.1.6. Модифицированный критерий знаков Мак-Нимара

Предположим, что над одной и той же группой объектов производятся два эксперимента и необходимо установить — меняется ли распределение частот от одного эксперимента к другому. В этом случае мы имеем также таблицу сопряженности 2×2, однако составляющие ее данные, вообще говоря, уже не будут независимыми.

Мак-Нимар [577] предложил простой критерий проверки изменения соотношения частот в таблице при изменении условий опыта. Пусть мы имеем таблицу:

Опыт 1	Опыт 2	
	+	−
+	a	b
−	c	d

Из нее следует, что значения a и d соответствуют неизменным условиям опыта (a — когда и опыт 1 и опыт 2 действуют на объект, d — когда ни один из опытов не действует на объект). Значения b и c соответствуют условиям, когда действует только один из опытов. Если $b \approx c$, то, следовательно, опыты не оказывают влияния на объект.

Для проверки равенства $b = c$ Мак-Нимар предложил критерий [577] $\chi^2 = \dfrac{(b-c)^2}{b+c+1}$, а при $b + c < 30$: $\chi^2 = \dfrac{(|b-c|-1)^2}{b+c+1}$.

Если $\chi^2 > \chi^2_\alpha(1)$, разница между b и c признается значимой ($\chi^2_\alpha(1)$ — α-квантиль распределения хи-квадрат с $f = 1$ степенью свободы).

Задача 350. *Предположим, 30 пациентов начали применять два препарата. При этом 8 пациентов признали сильным действие как первого, так и второго препаратов; 11 — слабым действие обоих препаратов; 16 признали действие первого препарата сильным, а второго слабым; 5 пациентов — наоборот. Необходимо проверить критерием Мак-Нимара при достоверности $\alpha = 0{,}95$ гипотезу о различии в действии препарата.*

Имеем таблицу:

Препарат 1	Препарат 2	
	сильное	слабое
сильное	(a) 8	16 (b)
слабое	(c) 5	11 (d)

Разница между препаратами проявляется в разнице частот (b) и (c), так как именно они фиксируют впечатления пациентов, связанные с различием воздействия исследуемых препаратов. Имеем

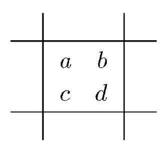
$$\chi^2 = \frac{(|b-c|-1)^2}{b+c+1} = \frac{(16-5-1)^2}{16+5+1} = 4{,}54.$$

Так как $\chi^2_{0,95}(1) = 3{,}84$ (см. табл. 55), а $\chi^2 = 4{,}54 > \chi^2_{0,95}$, то с вероятностью 0,95 следует признать значимой разницу в воздействии препаратов на различных пациентов.

5.2.2.4.1.1.7. G-критерий Вулфа

Критерий Вулфа [578] используется вместо критерия χ^2 для четыреклеточных таблиц (см. раздел 5.2.2.4.1.1.3). Он позволяет уменьшить количество вычислений и лучше обоснован теоретически, чем χ^2-критерий [9].

Критерий строится следующим образом. В таблице



a	b
c	d

сначала наименьшая из величин увеличивается на 1/2 (так называемая поправка Йэйтса). Затем соответствующим образом меняются все остальные значения в таблице с тем, чтобы величины $a+b$, $c+d$, $a+c$ и $b+d$ не менялись. Тогда таблица принимает следующий вид (считаем, что d является минимальным значением):

$a+\frac{1}{2}$	$b-\frac{1}{2}$
$c-\frac{1}{2}$	$d+\frac{1}{2}$

Статистика критерия Вулфа равна

$$G = 2\Bigg\{\left(a+\frac{1}{2}\right)\ln\left(a+\frac{1}{2}\right)+\left(b-\frac{1}{2}\right)\ln\left(b-\frac{1}{2}\right)+\left(c-\frac{1}{2}\right)\ln\left(c-\frac{1}{2}\right)+$$
$$+\left(d+\frac{1}{2}\right)\ln\left(d+\frac{1}{2}\right)-(a+b)\ln(a+b)-(c+d)\ln(c+d)-$$
$$-(a+c)\ln(a+c)-(b+d)\ln(b+d)+(a+b+c+d)\ln(a+b+c+d)\Bigg\}.$$

Если $G > \chi^2_\alpha(1)$, то связь признаков в таблице признается значимой.

Задача 351. *Для данных задачи* 345 *проверить сопряженность признаков в таблице критерием Вулфа при* $\alpha = 0{,}95$.

Имеем таблицу:

$95-1/2$	$15+1/2$	110
$70+1/2$	$40-1/2$	110
165	55	220

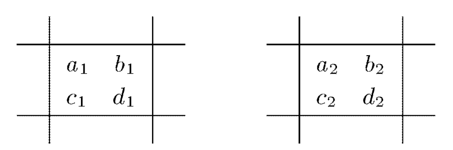
Вычисляем

$$G = 2\cdot\{94{,}5\cdot\ln 94{,}5+15{,}5\cdot\ln 15{,}5+70\cdot\ln 70+39{,}5\cdot\ln 39{,}5-$$
$$-110\cdot\ln 110-110\cdot\ln 110-165\cdot\ln 165-55\cdot\ln 55+220\cdot\ln 220\} = 14{,}337.$$

Из табл. 55 имеем $\chi^2_{0,95}(1) = 3{,}85$. Так как $G = 14{,}337 > \chi^2_{0,95}(1) = 3{,}85$, следует признать связь частот в таблице значимой.

5.2.2.4.1.1.8. Критерий Ле Роя для сравнения двух таблиц сопряженности 2×2

Предположим, что мы имеем в своем распоряжении две таблицы данных 2×2:



	a_1	b_1	
	c_1	d_1	

	a_2	b_2	
	c_2	d_2	

Необходимо проверить гипотезу о том, что обе таблицы статистически не различимы и являются выборками из единой совокупности.

Ле Рой [579] предложил для проверки такой гипотезы критерий, основанный на статистике

$$R = (a_1 + b_1 + c_1 + d_1 + a_2 + b_2 + c_2 + d_2) \times$$
$$\times \left\{ \begin{array}{l} \dfrac{1}{a_2+b_2+c_2+d_2}\left(\dfrac{a_2^2}{a_1+a_2} + \dfrac{b_2^2}{b_1+b_2} + \dfrac{c_2^2}{c_1+c_2} + \dfrac{d_2^2}{d_1+d_2}\right) + \\ + \dfrac{1}{a_1+b_1+c_1+d_1}\left(\dfrac{a_1^2}{a_1+a} + \dfrac{b_1^2}{b_1+b} + \dfrac{c_1^2}{c_1+c_2} + \dfrac{d_1^2}{d_1+d_2}\right) - 1 \end{array} \right\}.$$

Гипотеза о статистической неразличимости таблиц отклоняется с достоверностью α, если $R > \chi^2_\alpha(3)$, ($\chi^2_\alpha(3)$ — α-квантиль χ^2-распределения с $f = 3$ степенями свободы). Для употребляемых значений α укажем

$$\chi^2_{0.90}(3) = 6{,}251; \quad \chi^2_{0.95}(3) = 7{,}815; \quad \chi^2_{0.99}(3) = 11{,}345.$$

Следует помнить, что R-критерий применим, если все числа в таблицах превышают 3.

Задача 352. *Проверить гипотезу о статистической неразличимости двух таблиц сопряженности 2×2 критерием Ле Роя при доверительной вероятности $\alpha = 0{,}90$:*

			$\sum$
	15	48	63
	21	71	95
$\sum$	36	122	158

			$\sum$
	27	94	121
	112	456	568
$\sum$	139	550	689

Имеем $a_1 = 15$, $b_1 = 48$, $c_1 = 21$, $d_1 = 74$, $a_2 = 27$, $b_2 = 94$, $c_2 = 112$, $d_2 = 456$. Вычисляем

$$R = (15 + 48 + 21 + 74 + 27 + 94 + 112 + 456) \times$$
$$\times \left\{ \frac{1}{27+94+112+456} \cdot \left(\frac{27^2}{15+27} + \frac{94^2}{48+94} + \frac{112^2}{21+112} + \frac{456^2}{74+456}\right) + \right.$$
$$\left. + \frac{1}{15+48+21+74} \cdot \left(\frac{15^2}{15+27} + \frac{48^2}{48+94} + \frac{21^2}{21+112} + \frac{74^2}{74+456}\right) - 1 \right\} = 37{,}938.$$

Так как $R = 37{,}938 > \chi^2_{0.90}(3) = 6{,}251$, следует признать, что таблицы статистически различимы.

5.2.2.4.1.1.9. Выбор числа наблюдений для анализа таблиц сопряженности 2×2

Сформулируем следующую задачу. Необходимо для заданных допустимых уровней ошибок — первого рода α (вероятность отклонить верную гипотезу) и второго

рода β (вероятность принять неверную гипотезу) и заданных значениях $p_1 = \frac{a}{b}$ и $p_2 = \frac{c}{d}$ при $a + b = c + d = n$ определить необходимое число наблюдений — n, которое требуется провести, чтобы сравнить относительные частоты p_1 и p_2 между собой.

Необходимое число наблюдений n рассчитывается по формуле [9] (берется ближайшее целое число): $n = \dfrac{K}{(\arcsin\sqrt{p_1} - \arcsin\sqrt{p_2})^2}$, где K — коэффициент, приведенный в табл. 229.

Таблица 229

Значения коэффициента *K* [9]

α	β		
	0,2	0,1	0,01
0,05	12885	17250	30161
0,01	16474	21369	35537
0,001	19172	24426	43945

Задача 353. *Найти объем выборки, позволяющий сравнить относительные частоты в таблице сопряженности признаков* 2×2: $p_1 = 0{,}61$ *и* $p_2 = 0{,}38$ *при уровнях ошибки первого рода* $\alpha = 0{,}05$ *и второго рода* $\beta = 0{,}1$.

Из табл. 229 имеем $K = 17250$.

Тогда

$$n = \frac{17250}{\left(\arcsin\sqrt{0{,}61} - \arcsin\sqrt{0{,}38}\right)^2} = \frac{17250}{(51{,}3545 - 38{,}0567)^2} = 97{,}55.$$

Таким образом, необходимо иметь 98 наблюдений. Проверка разницы в таких таблицах обеспечивает мощность критерия χ^2, равную $1 - \beta = 0{,}90$.

5.2.2.4.2. Оценка связи признаков в многоклеточных таблицах сопряженности $r \times c$

Если результаты наблюдений могут быть классифицированы по трем или более качественным признакам, рассматриваются так называемые таблицы сопряженности $r \times c$ [132]:

					$\sum$
	n_{11}	n_{12}	$\ldots$	n_{1c}	n_1
	n_{21}	n_{22}	$\ldots$	n_{2c}	n_2
	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	n_{r1}	n_{r2}	$\ldots$	n_{rc}	n_r
$\sum$	n_1^*	n_2^*	$\ldots$	n_c^*	n

Здесь n_{ij} — число результатов наблюдений из общего числа n, обладающих признаками i и j одновременно. В качестве меры связи между признаками r и c (либо между r и c градациями двух признаков A и B) используется статистика $\chi^2 = n\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{n_{ij}^2}{n_i n_j^*} - 1\right)$, имеющая при независимости признаков

χ^2-распределение с $f = (r-1)(c-1)$ степенями свободы. Следовательно, если $\chi^2 > \chi^2_\alpha[(r-1)(c-1)]$, то с вероятностью α признается зависимость между изучаемыми признаками. Однако статистика χ^2 неудобна при оценке связи признаков, так как ее значения не нормированы и при $n \to \infty$ $\chi^2 \to \infty$ (в отличие от рассмотренных ранее коэффициентов корреляции, значения которых при любых n заключены между -1 и $+1$). Поэтому для оценки связей в таблицах $r \times c$ используются специальные коэффициенты сопряженности, предложенные Пирсоном и Чупровым [132]. Пирсон предложил коэффициент сопряженности в виде $K_P = \left(\dfrac{\chi^2}{n+\chi^2}\right)^{\frac{1}{2}}$. Значения K_P зависят от числа изучаемых признаков, что не позволяет использовать его для сравнения связей в таблицах с различными значениями r и c. Этого недостатка лишен коэффициент сопряженности Чупрова [132]

$$K_R = \left\{\frac{\chi^2}{n\left[(r-1)(c-1)\right]^{\frac{1}{2}}}\right\}^{\frac{1}{2}}.$$

Коэффициент K_R становится равным ± 1 в случае полной связи признаков только при $r = c$. Так, как коэффициенты сопряженности выражаются через χ^2, то проверка их значимости может быть осуществлена с помощью критических значений χ^2-распределения.

Если $K_P > K_P(\alpha)$ или $K_R > K_R(\alpha)$, то связь признаков признается существенной.

Задача 354. *Предположим, что в результате проверки партии электронных ламп трех типов (по 100 шт. каждого типа), изготовленных на пяти заводах, получены следующие количества годных ламп:*

Тип лампы	Завод-изготовитель					$\sum$
	1	2	3	4	5	
1	70	60	20	40	30	220
2	80	90	100	90	70	430
3	30	40	30	20	50	170
$\sum$	180	190	150	150	150	820

Необходимо проверить гипотезу о наличии связи между качеством ламп различного типа и заводом-изготовителем при доверительной вероятности $\alpha = 0{,}95$.

Имеем $r = 3$, $c = 5$, $(r-1)\cdot(c-1) = 8$. Тогда

$$\chi^2 = 820\left(\sum_{i=1}^{3}\sum_{j=1}^{5}\frac{n_{ij}^2}{n_i \cdot n_j^*} - 1\right) =$$

$$= 820\cdot\left(\frac{70^2}{180\cdot 220} + \frac{60^2}{190\cdot 220} + \frac{90^2}{150\cdot 220} + \ldots + \frac{20^2}{150\cdot 170} + \frac{50^2}{150\cdot 170} - 1\right) = 51{,}244.$$

Имеем из табл. 55: $\chi^2_{0,95}(8) = 15{,}507$. Так как $\chi^2 = 51{,}244 > \chi^2_{0,95}(8) = 15{,}507$, гипотеза о наличии связи между признаками подтверждается.

Вычислим теперь коэффициенты сопряженности:

$$K_p = \sqrt{\frac{51{,}244}{820 + 51{,}244}} = 0{,}242; \quad K_R = \sqrt{\frac{51{,}244}{820\cdot\sqrt{2\cdot 4}}} = 0{,}149.$$

5.3. Регрессионный анализ

Рассмотренные ранее методы дисперсионного и корреляционного анализа позволяют выявить наличие связи между случайными величинами и оценить силу этой связи.

Следующей ступенью является выявление конкретного функционального вида связи между случайными величинами.

При наличии корреляционной связи между x и y имеет место соотношение $F(y) = F(x, y)$, т. е. функция распределения случайной величины y зависит от значения случайной величины x. Любая функция распределения полностью определяется своими параметрами. Изменение функции распределения случайной величины y от x можно задать зависимостями

$$\mu_{1y} = f(x)\,; \quad \mu_{2y} = \sigma^2 = f_2(x)\,; \quad \mu_{3y} = f_3(x)\,; \quad \mu_{4y} = f_4\,(x)\,,$$

называемыми соответственно-регрессионной, скедастической, клитической и синагической. На практике обычно предполагается, что дисперсия и моменты высших порядков распределения y не зависят от значения x. Наибольший практический интерес представляет определение зависимости $\mu_{1y} = f(x)$, описывающей истинную зависимость между y и x. Зависимость средних значений $y(\mu_y)$ называется регрессией y по x, а методы нахождения таких зависимостей и оценки их статистических свойств составляют содержание регрессионного анализа.

По выборочным данным можно найти только оценку истинной регрессии, содержащую ошибку, связанную со случайностью выборки.

В основе регрессионного анализа лежит принцип наименьших квадратов, в соответствии с которым в качестве уравнения регрессии $y = f(x)$ выбирается функция, доставляющая минимум сумме квадратов разностей $s = \sum\limits_{i=1}^{n} [y_i - f\,(x_i)]^2$. Как правило, вид функции $f(x)$ определяется заранее, а методом наименьших квадратов определяются ее коэффициенты, минимизирующие s. Количественной мерой рассеяния значений y_i вокруг регрессии $f(x)$ является дисперсия

$$D = \frac{1}{n-k}\sum_{i=1}^{n} [y_i - f\,(x_i)]^2\,,$$

где k — число коэффициентов, входящих в аналитическое выражение регрессии (например, если $f(x)$ — многочлен степени l, то $k = l + 1$).

В зависимости от вида уравнения регрессии $y = f(x)$ различают линейную ($f(x)$ — многочлен первой степени) и нелинейную ($f(x)$ — многочлен степени $\geqslant 2$) регрессии.

Вид функции $f(x)$ выбирается исходя из особенностей исследуемого явления (процесса), а также из общего графического анализа зависимости между y и x. Подробно выбор формы функциональной зависимости для регрессии рассмотрен в [9, 14, 580].

Чаще всего ограничиваются рассмотрением линейной регрессионной модели, а при нелинейной зависимости $y = f(x)$ используют различные линеаризующие преобразования переменных y и x. Наиболее распространенные из этих преобразований приведены в табл. 230.

Схема регрессионного анализа включает в себя последовательное решение следующих задач: нахождение выборочной оценки истинной регрессии; оценки статистической значимости выборочной регрессии в сравнении с безусловным разбросом значений y_i, характеризующимся дисперсией σ_y^2; определение доверительных областей, с заданной вероятностью включающих в себя истинную регрессию.

Таблица 230

Линеаризующие функциональные преобразования ($y^* = a^* + b^*x^*$)

Исходная зависимость $y = f(x)$	Преобразование переменных		Преобразование коэффициентов	
	y^*	x^*	a^*	b^*
$y = a + \frac{b}{x}$	y	$\frac{1}{x}$	a	b
$y = \frac{a}{b+x}$	$\frac{1}{y}$	x	$\frac{a}{b}$	$\frac{1}{a}$
$y = \frac{ax}{b+x}$	$\frac{1}{y}$	$\frac{1}{x}$	$\frac{b}{a}$	$\frac{1}{a}$
$y = \frac{x}{a+bx}$	$\frac{x}{y}$	x	a	b
$y = ab^x$	$\lg y$	x	$\lg a$	$\lg b$
$y = ax^b$	$\lg y$	$\lg x$	$\lg a$	b
$y = ae^{bx}$	$\ln y$	x	$\ln a$	b
$y = ae^{\frac{b}{x}}$	$\ln y$	1	$\ln a$	b
$y = a + bx^n$	y	x^n	a	b

Среди дополнительных задач, позволяющих получить полную статистическую картину изучаемой регрессии, отметим: анализ так называемых регрессионных остатков (разница между выборочной регрессией и выборочными значениями функции); анализ наличия грубых отклонений от регрессии (выбросов); построение толерантных границ для регрессии.

Эти задачи практически не используются в повседневной работе инженеров и исследователей, поэтому им далее уделяется необходимое внимание.

Разработанный в настоящее время аппарат регрессионного анализа предполагает, что значения y_i взаимно независимы и нормально распределены. Выполнение этих условий должно быть предварительно проверено с помощью критериев нормальности (см. раздел 3.2.2) и критериев сравнения дисперсий (см. раздел 4.1.1.4). Наиболее полное изложение прикладного регрессионного анализа содержится в [581].

5.3.1. Линейный регрессионный анализ

Линейный регрессионный анализ исходит из наличия зависимости $y = \alpha + \beta x$, где α и β — неизвестные коэффициенты регрессии. Выборочные оценки α и β в дальнейшем будем обозначать a и b соответственно.

5.3.1.1. Оценка коэффициентов регрессии

5.3.1.1.1. Оценка наименьших квадратов

Оценки наименьших квадратов являются решениями системы нормальных уравнений, строящихся по совокупности наблюдаемых значений y_i для совокупности значений x_i:

$$\begin{cases} \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} (\alpha + \beta x_i) = 0; \\ \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} (\alpha + \beta x_i)\, x_i = 0, \end{cases}$$

из которой следует система

$$\begin{cases} n\alpha + \beta \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i; \\ \alpha \sum_{i=1}^{n} x_i + \beta \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i. \end{cases}$$

Решение системы дает искомые оценки коэффициентов регрессии:

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}; \quad a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n}.$$

Для проверки правильности вычислений можно использовать соотношения

$$\bar{y} = a + b\bar{x}, \quad \text{где} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i; \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Вычисления a и b существенно упрощаются, если интервалы между значениями независимой переменной x постоянны, т. е., если $x_{i+1} - x_i = l = \mathrm{const}$ $(i = 1, 2, \ldots, n-1)$. Наиболее эффективен метод оценки, предложенный авторами в [582]. Пусть $\tilde{x}_i = \dfrac{x_i - x_1}{l} + 1$, т. е. последовательность $x_1, x_2, \ldots, x_n$ трансформируется непосредственно в последовательность $i = 1, 2, \ldots, n$ и зависимость $y = a + bx$ трансформируется в завсимость $y = \tilde{a} + \tilde{b}i$. Оценки для этого случая имеют вид

$$\tilde{b} = \frac{12\sum_{i=1}^{n} iy_i - 6(n+1)\sum_{i=1}^{n} y_i}{n(n^2-1)}; \quad \tilde{a} = \bar{y} - \tilde{b}\bar{x} = \frac{\sum_{i=1}^{n} y_i - \tilde{b}\sum_{i=1}^{n} i}{n}.$$

Дисперсии этих оценок равны: $\mathbf{D}(\tilde{b}) = \dfrac{1}{n-n}\sigma^2$; $\mathbf{D}(\tilde{a}) = 2\dfrac{2n+1}{n(n-1)}\sigma^2$, где σ^2 — дисперсия остатков.

Для вычисления оценок $\tilde{a}$ и $\tilde{b}$ необходимо вычисление сумм $\sum y_i$ и $\sum iy_i$, что требует n умножений и $2n$ сложений. Авторы работы [582] предложили более экономичный линейный метод оценивания, который при незначительной потере в точности позволяет существенно сократить время вычислений. Эти оценки имеют вид:

$$\tilde{b} = \frac{4{,}5}{n^2}(S_3 - S_1); \quad \tilde{a} = \frac{1}{n}\left(S_1 + S_2 + S_3 - \tilde{b}\frac{n(n-1)}{2}\right),$$

где $S_1 = \sum_{i=1}^{k} y_i$; $S_2 = \sum_{i=k+1}^{n-k} y_i$; $S_3 = \sum_{i=n-k+1}^{n} y_i$; k — ближайшее целое к $\dfrac{n}{3}$.

Расчет на ЭВМ этим способом занимает в $\approx 3{,}2$ раза меньше времени, а дисперсия оценок возрастает на 12,5% и 9,4% соответственно.

В заключение отметим характерную особенность регрессионных уравнений, о которой следует помнить инженерам и исследователям. Регрессия y по x: $y = \alpha + \beta x$ не эквивалентна в общем случае регрессии x по y: $-x = \alpha^* + \beta^* y$.

Если S_y и S_x — стандартные отклонения совокупностей значений y и x соответственно ($S_y^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$; $S_x^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$), то регрессии $y = f(x)$

и $x = \varphi(y)$ можно записать следующим образом:

$$y = \bar{y} + r\frac{S_y}{S_x}(x - \bar{x}); \quad x = \bar{x} + r\frac{S_x}{S_y}(y - \bar{y}),$$

где r — коэффициент корреляции.

Отсюда видно, что регрессии y по x и x по y совпадают только в одном случае, когда существует абсолютная корреляция между y и x, т. е. когда $|r| = 1$. При $r = 0$ прямые регрессии y по x и x по y перпендикулярны. Тогда $\beta = r\dfrac{S_y}{S_x}$; $\beta^* = r\dfrac{S_x}{S_y}$.

При $S_x = S_y$ коэффициенты корреляции и регрессии совпадают.

Задача 355. *В результате наблюдений за зависимостью $y = f(x)$ получены следующие данные:*

$$\begin{array}{lcccccccccc} y_i: & 2 & 3 & 7 & 10 & 11 & 13 & 18 & 21 & 25 & 31; \\ x_i: & 8 & 11 & 14 & 18 & 4 & 26 & 31 & 32 & 34 & 41. \end{array}$$

Необходимо найти оценку коэффициентов регрессии y по x методом наименьших квадратов.

Находим $\sum\limits_{i=1}^{10} x_i = 219$; $\sum\limits_{i=1}^{10} y_i = 141$; $\sum\limits_{i=1}^{10} x_i^2 = 6219$; $\sum\limits_{i=1}^{10} x_i y_i = 4060$.

Далее вычисляем оценки

$$b = \frac{10 \cdot 4060 - 219 \cdot 141}{10 \cdot 6219 - 219^2} = 0{,}68318; \quad a = \frac{141 - 0{,}68318 \cdot 219}{10} = -0{,}86164.$$

Следовательно, уравнение регрессии y по x имеет вид

$$y = -0{,}86164 + 0{,}68318 \cdot x.$$

Задача 356. *В результате наблюдения за зависимостью $y = f(x)$ получены следующие данные:*

$$\begin{array}{lccccccccccccccc} y_i: & 13 & 18 & 24 & 21 & 25 & 31 & 36 & 41 & 35 & 41 & 48 & 56 & 61 & 60 & 70; \\ x_i: & 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 & 22 & 24 & 26 & 28 & 30. \end{array}$$

Необходимо найти оценки коэффициентов регрессии y по x.

Введем замену переменных $\tilde{x}_i = \dfrac{x_i - x_1}{l} + 1 = \dfrac{x_i - 2}{2} + 1$.

Вычисляем оценки наименьших квадратов

$$\tilde{b} = \frac{12 \cdot \sum\limits_{i=1}^{15} i \cdot y_i - 6 \cdot (15+1) \cdot \sum\limits_{i=1}^{15} y_i}{15 \cdot (225 - 1)} = \frac{12 \cdot 5704 - 6 \cdot 16 \cdot 580}{15 \cdot 224} = 3{,}8;$$

$$\tilde{a} = \frac{580}{15} - \frac{3{,}8 \cdot (15+1)}{2} = 8{,}266.$$

Воспользуемся теперь линейными оценками. Примем $k = n/3 = 5$ и вычисляем

$$S_1 = \sum_{i=1}^{5} y_i = 101; \quad S_2 = \sum_{i=6}^{11} y_i = 184; \quad S_3 = \sum_{i=11}^{15} y_i = 295;$$

$$\tilde{b} = \frac{4{,}5}{15^2} \cdot (S_3 - S_1) = \frac{4{,}5 \cdot (295 - 101)}{225} = 3{,}88;$$

$$\tilde{a} = \frac{1}{15} \cdot \left(101 + 184 + 295 - 3{,}88 \cdot \frac{15 \cdot 16}{2}\right) = 7{,}626.$$

Видим, что при существенно меньшем количестве вычислений результат практически не отличается от предыдущего (естественно, что наибольший выигрыш во времени и наиболее точный результат будут достигнуты при больших объемах выборок).

5.3.1.1.2. Простейшие оценки коэффициентов регрессии

5.3.1.1.2.1. Метод Бартлетта–Кенуя

Пары наблюдений (y_i, x_i) упорядочиваются по x и разбиваются на 3 примерно равные группы (причем первая и последняя группы должны быть обязательно равного объема). В каждой группе находятся суммы $\sum y_i$ и $\sum x_i$ (обозначим их соответственно Y_1, Y_2, Y_3 и X_1, X_2, X_3).

Тогда коэффициент регрессии оценивается с помощью соотношения [121, 583, 584]:

$$\tilde{b} = \frac{Y_3 - Y_1}{X_3 - X_1} \quad \text{с ошибкой} \quad S_{\tilde{b}} = \frac{0{,}8s\sqrt{n}}{X_3 - X_1},$$

где $s = \dfrac{8}{9}\sum\limits_{i=1}^{n}\dfrac{|y_i - y_{i+1}|}{n}$; $\tilde{a} = y - \tilde{b}\bar{x}$.

Если n пар наблюдений разбиваются на четыре группы, содержащие в себе 1/6, 1/3, 1/3 и 1/6 часть наблюдений, то

$$\tilde{b} = \frac{3Y_1 + Y_2 - Y_3 - 3Y_4}{3X_1 + X_2 - X_3 - 3X_4} \quad \text{с ошибкой} \quad S_{\tilde{b}} = \frac{1{,}9s\sqrt{n}}{3X_1 + X_2 - X_3 - 3X_4}.$$

Эти оценки применимы для больших выборок при $n \geqslant 100$.

Задача 357. *Для данных задачи* 356 *найти оценку коэффициентов регрессии методом Бартлетта–Кенуя.*

Разбиваем пары наблюдений (y_i, x_i) на три части и находим соответствующие суммы для переменной $\tilde{x}_i = \dfrac{x_i}{2}$.

$$X_1 = 15, \quad X_2 = 40, \quad X_3 = 65, \quad Y_1 = 101, \quad Y_2 = 184, \quad Y_3 = 295.$$

Тогда имеем

$$\tilde{b} = \frac{295 - 101}{65 - 15} = 3{,}88,$$

что очень близко к полученному ранее в задаче 355 значению $\tilde{b} = 3{,}8$.

5.3.1.1.2.2. Метод Керрича

Для частного случая зависимости $y = bx$ $(a = 0)$ Керрич [585] предложил следующий простой метод оценки. Вычисляем разности

$$d_i = \lg y_i - \lg x_i; \quad \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad \text{и} \quad S_{\bar{d}} = \left\{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2\right\}^{\frac{1}{2}}.$$

Так как любое отношение y_i/x_i является оценкой b, то любое значение d_i является оценкой $\lg b$.

Когда $S_{\bar{d}}/\bar{d} \ll 1$, оценкой $\lg b$ является величина $\bar{d}$. Следовательно, оценка будет равна $\tilde{b} = 10^{\bar{d}}$.

Задача 358. *Для совокупности значений*

x_i:	1	2	3	4	5	6	7	8	9	10;
y_i:	3	8	6	16	15	18	21	29	28	32

найти оценку коэффициента регрессии методом Керрича.

Вычисляем последовательность значений $d_i = \lg y_i - \lg x_i$:

d_i: 0,477; 0,602; 0,301; 0,602; 0,477; 0,477; 0,559; 0,559; 0,493; 0,505.

Далее $\tilde{d} = 0{,}497$ и $\tilde{b} = 10^{0{,}497} = 3{,}14$, что близко к обычной оценке $\tilde{b} = 3{,}23$.

5.3.1.1.3. Робастные методы оценки параметров регрессии

Напомним, что одним из важнейших исходных требований регрессионного анализа является обязательная нормальность распределения наблюдаемых случайных величин. В связи с тем, что эта предпосылка не всегда выполняется, разработаны специальные методы, устойчивые (робастные) к отклонениям распределения исходных величин от нормального распределения. Среди них можно отметить оценки коэффициентов регрессии с помощью коэффициентов корреляции Кендалла τ [586], „повторных“ медиан [587], метода областей Дениэлса [588].

Мы рассмотрим два критерия, наиболее полно отображающих суть подходов к построению робастных методов оценки параметров регрессии.

5.3.1.1.3.1. Медианный критерий Брауна–Муда

Предположим, мы имеем n пар (n — четное) наблюдений (x_i, y_i). Разбиваем все наблюдения на две группы по значению x: группу значений, превышающих $\tilde{x}$, и группу значений $x_i < \tilde{x}$. Предположим, что мы располагаем априорными оценками коэффициентов регрессии a_0 и b_0. Для них можно найти регрессию $\hat{y}_i = a_0 + b_0 x_i$ и вычислить регрессионные остатки $\Delta y_i = y_i - \hat{y}_i$. Затем определим количества положительных ($\Delta y_i > 0$) остатков m_1 и m_2 в двух группах.

Статистикой критерия Брауна–Муда [589] является величина

$$A = \frac{S}{n}\left[\left(m_1 - \frac{n}{4}\right)^2 + \left(m_2 - \frac{n}{4}\right)^2\right].$$

Гипотеза адекватности регрессии, а следовательно, пригодность выбранных коэффициентов a_0 и b_0 отклоняется, если $A > \chi^2_\alpha$ (χ^2_α — α-квантиль χ^2-распределения с $f = 2$ степенями свободы). В ряде случаев более эффективной может быть модификация A-критерия Брауна–Муда

$$A^* = \frac{2}{\sqrt{n}}\left(\left|m_1 - \frac{n}{4}\right| + \left|m_2 - \frac{n}{4}\right|\right).$$

Критические значения A^*-статистики равны $A^*_{0,95} = 2{,}237$, $A^*_{0,99} = 2{,}806$ [589]. При $A^* > A^*_\alpha$ гипотеза о пригодности a_0 и b_0 отклоняется с вероятностью α. В этом случае оценки a_0 и b_0 заменяются на другие, и итерации продолжаются до тех пор, пока критерии не будут отклоняться критическими значениями.

Задача 359. *При условиях задачи 358 провести оценку коэффициентов регрессии с помощью критериев Брауна–Муда* ($\alpha = 0{,}95$).

Выберем оценки $a_0 = 1$ и $b_0 = 3$ и вычислим

$$\hat{y}_i = a_0 + b \cdot x_i \to 4,\ 7,\ 10,\ 13,\ 16,\ 19,\ 22,\ 25,\ 28,\ 31;$$
$$\Delta y_i = y_i - \hat{y}_i \to -1,\ +1,\ -4,\ +3,\ -1,\ -1,\ -1,\ +4,\ 0,\ -1.$$

Медиана равна $\tilde{x} = \dfrac{5+6}{2} = 6{,}5$. Для $x_i > \tilde{x}$ имеем $m_1 = 2$ значения $\Delta y > 0$ (это 4 и 1). Для $x_i < \tilde{x}$ имеем $m_2 = 2$ (это 1 и 3). Тогда

$$A = \frac{8}{10}\cdot\left[\left(2 - \frac{10}{4}\right)^2 + \left(2 - \frac{10}{4}\right)^2\right] = 0{,}4;$$
$$A^* = \frac{2}{\sqrt{10}}\cdot\left(\left|2 - \frac{10}{4}\right| + \left|2 - \frac{10}{4}\right|\right) = 0{,}632.$$

Так как $A = 0{,}4 < \chi^2_{0,95}(2) = 6$ и $A^* = 0{,}632 < A^*_{0,95} = 2{,}237$, принятые оценки $a_0 = 1$ и $b_0 = 3$ не отклоняются критерием Брауна–Муда.

5.3.1.1.3.2. Оценка Тейла

По n парам (x_i, y_i), где среди значений x_i нет повторяющихся, построим $N = n(n-1)$ пар оценок угловых коэффициентов

$$\beta_{ij} = \frac{y_j - y_i}{x_j - x_i}, \quad i < j.$$

Оценкой коэффициента регрессии β является медиана [590] $b = \operatorname{med}(\beta_{ij})$.

Так как N — четное , то $b = \frac{1}{2}\left(\beta_{ij}^{k} + \beta_{ij}^{k+1}\right)$, где $k = \frac{N}{2}$.

Двусторонний $\alpha \cdot 100\%$-й доверительный интервал для оценки Тейла имеет вид [590]

$$\beta_{ij}\left(\frac{N - C_\alpha}{2}\right) \leqslant b \leqslant \beta_{ij}\left(\frac{N + C_\alpha}{2}\right),$$

где $C_\alpha = u_{\frac{1+\alpha}{2}}\sqrt{\frac{n(n-1)(2n+5)}{18}}$ (при > 10); u_γ — γ-квантиль стандартного нормального распределения.

Задача 360. *В условиях задачи 358 найти оценку Тейла для регрессионного коэффициента β и вычислить его 95%-й двусторонний доверительный интервал.*

Найдем $\frac{n \cdot (n-1)}{2} = \frac{10 \cdot 9}{2} = 45$ пар значений (x_i, x_j) для различных i:

$$(x_1, x_2),\ (x_1, x_3),\ \ldots,\ (x_1, x_{10}),\ (x_2, x_3),\ \ldots,\ (x_2, x_{10}),\ \ldots,$$
$$(x_7, x_8),\ (x_7, x_9),\ (x_7, x_{10}),\ (x_8, x_9),\ (x_8, x_{10}),\ (x_9, x_{10}).$$

Вычисляем 45 оценок $\beta_{ij} (i < j)$ вида $\beta_{ij} = \frac{y_j - y_i}{x_j - x_i}$. Результаты вычислений сводим в таблицу:

i	j									
	1	2	3	4	5	6	7	8	9	10
1	–	5	3/2	13/3	3	3	3	26/7	25/8	29/9
2	–	–	2	4	7/3	10/4	13/5	21/6	20/8	24/9
3	–	–	–	10	9/2	4	15	23/5	22/6	26/7
4	–	–	–	–	1	1	5/3	13/4	12/5	16/6
5	–	–	–	–	–	3	3	14/3	13/4	17/5
6	–	–	–	–	–	–	3	11/2	10/3	14/4
7	–	–	–	–	–	–	–	8	7/2	11/3
8	–	–	–	–	–	–	–	–	1	3/2
9	–	–	–	–	–	–	–	–	–	4

Ранжированный по возрастанию ряд значений β_{ij} имеет вид

$$-2,\ -1,\ -1,\ \frac{3}{2},\ \frac{3}{2},\ \frac{5}{3},\ \frac{7}{3},\ \frac{12}{5},\ \frac{5}{2},\ \frac{5}{2},\ \frac{5}{2},\ \frac{13}{5},\ \frac{13}{5},\ \frac{24}{9},\ 3,\ 3,\ 3,\ 3,\ 3,\ 3,\ \frac{25}{8},\ \frac{29}{9},\ \frac{13}{4},$$
$$\frac{13}{4},\ \frac{10}{3},\ \frac{17}{5},\ \frac{7}{2},\ \frac{7}{2},\ \frac{7}{2},\ \frac{11}{3},\ \frac{22}{6},\ \frac{22}{6},\ \frac{22}{6},\ \frac{15}{4},\ 4,\ 4,\ 4,\ \frac{13}{3},\ \frac{9}{2},\ \frac{14}{3},\ \frac{14}{3},\ 5,\ \frac{11}{2},\ 8,\ 10.$$

Медиана этого ряда равна 23-му порядковому значению, т. е. $b = \operatorname{med}(\beta_{ij}) = \beta_{23} = \frac{13}{4} = 3{,}25$.

Для двусторонней оценки при $\alpha = 0{,}95$ получаем

$$C_{0,95} = u_{0,975} \cdot \sqrt{\frac{10 \cdot 9 \cdot 25}{18}} = 1{,}96 \cdot 11{,}18 = 21{,}91.$$

Находим β с номерами $\dfrac{N - C_\alpha}{2} = \dfrac{45 - 21{,}91}{2} = 11$ (ближайшее меньшее целое число) и $\dfrac{N + C_\alpha}{2} = \dfrac{45 + 21,91}{2} = 34$ (ближайшее большее целое число).

Имеем $\beta_{11} = 2{,}5$ и $\beta_{34} = 3{,}75$. Следовательно, искомый доверительный интервал для коэффициента регрессии равен $2{,}5 \leqslant b \leqslant 3{,}75$.

5.3.1.2. Статистическое оценивание регрессии

Статистическое исследование регрессии включает в себя:

— проверку гипотез о значениях коэффициентов регрессии (т. е. о статистической неразличимости выборочных значений y_i и $\hat{y}_i$, вычисляемых по уравнению регрессии);

— построение доверительных интервалов для коэффициентов регрессии и доверительных областей для регрессии;

— анализ статистической однородности регрессии (отсутствие выбросов);

— анализ независимости регрессионных остатков.

Ниже рассматриваются различные методы решения перечисленных задач.

5.3.1.2.1. Статистический анализ коэффициентов регрессии

5.3.1.2.1.1. Оценки наименьших квадратов

Статистические выводы относительно коэффициента β регрессии $y = \alpha + \beta x$ могут быть получены с помощью статистики

$$t_\beta = \frac{b - \beta}{S_\beta},$$

где

$$S_\beta = \frac{S}{S_x\sqrt{n-1}}; \quad S^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - a - bx_i\right)^2;$$

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2; \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i;$$

β — истинное значение коэффициента регрессии; b — выборочная оценка коэффициента регрессии.

Статистика t_β при справедливости нулевой гипотезы $H_0\colon \beta = b$ имеет распределение Стьюдента с $f = n - 2$ степенями свободы.

Следовательно, с помощью квантилей распределения Стьюдента можно проверить гипотезу равенства β заданному значению, гипотезу о значимости коэффициента регрессии (существенности его отклонения от нуля), построить доверительный интервал для коэффициента β. Значение коэффициента β регрессии является значимым с достоверностью α, если $|b| > t_{\frac{1+\alpha}{2}} S_\beta$.

Гипотеза о равенстве коэффициента β заданному значению β_0 принимается, если $|\beta - b| < t_{\frac{1+\alpha}{2}} S_\beta$.

И, наконец, двусторонний $\alpha \cdot 100\%$-й доверительный интервал для β имеет вид

$$b - S_\beta t_{\frac{1+\alpha}{2}} \leqslant \beta \leqslant b + s_\beta t_{\frac{1+\alpha}{2}}.$$

Статистические выводы относительно коэффициента α могут быть получены с помощью статистики

$$t_\alpha = \frac{a-\alpha}{S_\alpha}, \quad \text{где} \quad S_\alpha = S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{(n-1)\,S_x^2}},$$

a, α — соответственно выборочная оценка и истинное значение коэффициента α; S и S_x определены выше для t_β.

При $H_0\colon a = \alpha$ статистика t_α имеет распределение Стьюдента с $f = n-2$ степенями свободы. Проверка гипотез о значениях коэффициента α и построение доверительных интервалов для него выполняются по аналогии с коэффициентом β.

Задача 361. *Для совокупности данных*

x_i:	1,2	2,4	2,8	4,2	5,9	6,8	8,1	9,2	10,1	11,0;
y_i:	7	12	17	24	29	38	46	45	54	68

найти оценки коэффициентов α и β регрессии $y = \alpha + \beta x$ и провести их статистический анализ при доверительной вероятности $\alpha = 0,95$ (не путать с α — коэффициентом регрессии!).

Вычисляем оценку β (см. раздел 5.3.1.1.1): $b = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$.

Имеем

$$\sum_{i=1}^{10} x_i = 61{,}7; \quad \left(\sum_{i=1}^{10} x_i\right)^2 = 3806{,}89; \quad \sum_{i=1}^{10} x_i^2 = 486{,}99; \quad \sum_{i=1}^{10} y_i = 340; \quad \sum_{i=1}^{10} x_i y_i = 2695{,}1.$$

Тогда

$$b = \frac{10\cdot 2695{,}1 - 61{,}7\cdot 340}{10\cdot 486{,}99 - 3806{,}89} = 5{,}6189; \quad a = \frac{\sum\limits_{i=1}^{10} y_i - b\cdot\sum\limits_{i=1}^{10} x_i}{n} = \frac{340 - 5{,}6189\cdot 61{,}7}{10} = -0{,}668.$$

Проверим теперь значимость полученных коэффициентов (существенность их отклонения от нуля). Вычислим предварительно

$$\bar{x} = 6{,}17; \quad S_x^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2 = 11{,}8112; \quad (S_x = 3{,}3467).$$

Далее вычисляем значение дисперсии $S^2 = \dfrac{1}{n-2}\sum\limits_{i=1}^{n}(y_i - \hat{y}_i)^2$, где $\hat{y}_i = a + b\cdot x_i$.

В нашем случае

$\hat{y}_i \to 6{,}075;\ 12{,}818;\ 15{,}065;\ 22{,}932;\ 32{,}484;\ 37{,}541;\ 44{,}846;\ 51{,}027;\ 56{,}084;\ 61{,}141.$

Вычисляем далее

$$S^2 = \frac{1}{8}\cdot\sum_{i=1}^{10}(y_i - \hat{y}_i)^2 =$$
$$= 0{,}125\cdot\left[(7-6{,}075)^2 + (12-12{,}818)^2 + \ldots + (68-61{,}141)^2\right] = 13{,}4755;$$

$$S_\beta = \frac{S}{S_x\cdot\sqrt{n-1}} = \frac{\sqrt{13{,}4755}}{3{,}3467\cdot 3} = 0{,}3656;$$

$$S_\alpha = S\cdot\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{(n-1)\cdot S_x^2}} = 3{,}671\cdot\sqrt{\frac{1}{10} + \frac{6{,}17^2}{9\cdot 11{,}8112}} = 2{,}486.$$

Для уровня достоверности $\alpha = 0{,}95$ имеем $t_{\frac{1-0{,}95}{2}}(n-2) = t_{0{,}975}(8) = 2{,}306$.

Проверяем значимость коэффициента β:

$$|b| = 5{,}1689 > t_{0,975}(8) \cdot S_\beta = 2{,}306 \cdot 0{,}3656 = 0{,}843,$$

следовательно, с достоверностью 0,95 делаем вывод о значимости коэффициента регрессии.

Проверяем гипотезу $H\colon \beta = \beta_0 = 5$ (о равенстве коэффициента регрессии $\beta_0 = 5$):

$$|5{,}619 - 5| = 0{,}619 < t_{0,975}(8) \cdot S_\beta = 0{,}843,$$

т. е. гипотеза о равенстве $\beta = 5$ не отклоняется.

И, наконец, доверительный интервал для β равен

$$5{,}9 - 2{,}306 \cdot 0{,}3656 = 4{,}776 \leqslant \beta \leqslant 6{,}462 = 5{,}619 + 2{,}306 \cdot 0{,}3656.$$

Аналогичные задачи решаем теперь для коэффициента α. Проверим гипотезу

$$H_0\colon \alpha = 0\, |a| = 0{,}668 < t_{0,975} \cdot S_\alpha = 2{,}306 \cdot 2{,}485 = 5{,}73.$$

Следовательно, коэффициент α с вероятностью 0,95 не отличается значимо от нуля, т. е. его значение может быть приравнено к нулю.

Двусторонний доверительный интервал для α имееет вид

$$\hat{a} - t_{0,975} \cdot S_\alpha \leqslant a \leqslant \hat{a} + t_{0,975} \cdot S_\alpha;$$
$$-0{,}668 - 2{,}306 \cdot 2{,}485 = -6{,}398 \leqslant a \leqslant 5{,}602 = -0{,}668 + 2{,}306 \cdot 2{,}485.$$

Таким образом, уравнение регрессии y по x адекватно отображается уравнением $y = 5{,}619 \cdot x$.

5.3.1.2.1.2. Робастные оценки Тейла

Для робастных оценок β Тейла (см. раздел 5.3.1.1.3.2) проверку гипотезы $H_0\colon \beta = \beta_0$ о равенстве углового коэффициента регрессии β заданному значению β_0 проводим следующим образом. Вычисляем разности $\Delta y_i = y_i - \beta_0 x_i$ $(i = 1, 2, \ldots, n)$. Статистикой критерия для проверки нулевой гипотезы является величина

$$C = \sum_{i<j}^{n} \delta(\Delta y_j - \Delta y_i)\,, \quad \text{где} \quad \delta(x) = \begin{cases} 1, & \text{если } x > 0; \\ 0, & \text{если } x = 0; \\ -1, & \text{если } x < 0. \end{cases}$$

Гипотеза $H_0\colon \beta = \beta_0$ отклоняется с достоверностью α, если (при $n > 10$) [18] $|c| > u_{\frac{1+\alpha}{2}} \sqrt{\dfrac{n\,(n-1)\,(2n+5)}{18}}$.

Построение доверительного интервала для β рассмотрено в разделе 5.3.1.1.3.2 (см. задачу 360).

Задача 362. *Для данных задачи 358 проверить гипотезу о равенстве нулю коэффициента регрессии $H_0\colon \beta = 0$ (т. е. гипотезу о значимости регрессии) при доверительной вероятности $\alpha = 0{,}95$.*

В нашем случае n разностей Δy_i совпадают с n значениями y_i. Имеем

$$C = \sum_{i<j}^{10} (y_j - y_i) = 9 - 1 + 7 + 7 - 1 + 5 + 5 + 4 + 3 - 1 + 1 + 1 = 39;$$

$$u_{\frac{1+\alpha}{2}} = u_{0,975} = 1{,}96; \qquad \sqrt{\frac{n(n-1)(2n+5)}{18}} = 11{,}180.$$

Тогда $|c| = 39 > 1{,}96 \cdot 11{,}180 = 21{,}913$, т. е. значение коэффициента β следует признать отличным от нуля.

5.3.1.2.2. Статистический анализ уравнения регрессии

Целью статистического анализа уравнения регрессии является установление его адекватности наблюдаемым экспериментальным данным. Под адекватностью уравнения регрессии понимается статистическая неразличимость результатов вычислений по уравнению регрессии и наблюдаемых случайных величин.

5.3.1.2.2.1. Оценка адекватности регрессии

Количественной мерой адекватности является отношение дисперсии S^2, определяемой рассеянием значений y_i вокруг линии регрессии, к дисперсии S_y^2 естественного рассеяния значений y_i вокруг своих средних $\bar{y}_i$. На привычном для инженера и исследователя языке это можно сформулировать так: ошибки, обусловленные заменой истинной зависимости на выборочную регрессию, находятся на уровне естественного разброса наблюдаемых случайных величин.

Если $\dfrac{S^2}{S_y^2} > F_\alpha$, где F_α — α-квантиль распределения Фишера с $f_1 = n-2$ и $f_2 = m-1$ степенями свободы, то ошибка в определении регрессии с доверительной вероятностью α признается статистически значимой (m — объем выборки, по которой выполнена оценка дисперсии S_y^2, т. е. число дублируемых наблюдений для каждой серии y_i).

Если дисперсия S_y^2 известна заранее (что бывает весьма редко), то $f_2 \to \infty$. Если дисперсия S_y^2 определяется по дублируемым значениям y_i, то ее оценкой является средневзвешенная дисперсия

$$S_y^2 = \frac{1}{n}\sum_{i=1}^{n} S_i^2, \quad \text{где} \quad S_i^2 = \frac{1}{m-1}\sum_{j=1}^{m}(y_{ij}-\bar{y}_i)^2; \quad \bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_j.$$

Напомним, что $S^2 = \dfrac{1}{n-2}\displaystyle\sum_{i=1}^{n}(y_i - a - bx_i)^2$.

Задача 363. *Проверить адекватность регрессии для данных задачи 361 при доверительной вероятности $\alpha = 0{,}95$, если для оценки S_y^2 предварительно проводилась серия наблюдений над случайной величиной y при неизменной величине x ($m = 10$):*

$$y_{ij}: \quad 12 \quad 14 \quad 11 \quad 10 \quad 8 \quad 6 \quad 7 \quad 15 \quad 13.$$

Ранее (см. задачу 361) мы получили $S^2 = 13{,}4755$. По отдельной серии наблюдений находим оценку $S_y^2 = \dfrac{1}{m-1}\cdot\displaystyle\sum_{i=1}^{m}(y_{ij}-\bar{y}_i)^2 = 11{,}733$ ($S_y = 3{,}425$).

Далее имеем $F = \dfrac{S^2}{S_y^2} = \dfrac{13{,}4755}{11{,}733} = 1{,}148$. Из таблиц F-распределения находим $F_{0,95}(f_1 = n-2 = 8;\ f_2 = m-2 = 8) = 3{,}438$.

Так как $F = 1{,}148 < F_{0,95}(8{,}8) = 3{,}438$, с вероятностью $\alpha = 0{,}95$ следует сделать вывод о статистической неразличимости сравниваемых дисперсий, а следовательно, об адекватности уравнения регрессии.

5.3.1.2.2.2. Анализ регрессионных остатков

Определенную информацию об адекватности уравнения регрессии дает исследование остатков вида $e_i = y_i - \hat{y}_i$, где $\hat{y}_i = a + bx_i$. Если выборочная регрессия $\hat{y}$ удовлетворительно описывает истинную зависимость между y и x, то остатки e_i должны быть независимыми нормально распределенными случайными величинами

с нулевым средним и в значениях e_i должен отсутствовать тренд. Нормальность распределения остатков e_i может быть установлена одним из критериев согласия (см. раздел 3.2.2).

Гипотезу о равенстве $M(e) = 0$ можно проверить любым параметрическим или непараметрическим критерием сравнения среднего с заданным значением (в нашем случае с нулем), изложенным в главе 4; гипотезу об отсутствии тренда — одним из критериев тренда и случайности (см. раздел 4.3).

Независимость в последовательности значений e_i $(i = 1, \ldots, n)$ может быть проверена с помощью сериального коэффициента корреляции Дарбина–Ватсона [591]. Статистика сериального коэффициента корреляции Дарбина–Ватсона имеет вид

$$D = \frac{\sum\limits_{i=2}^{n} (e_i - e_{i-1})^2}{\sum\limits_{i=1}^{n} e_i^2}.$$

Если $D > D_1(\alpha)$ или $D > 4 - D_1(\alpha)$, то с достоверностью α принимается гипотеза о наличии соответственно отрицательной или положительной корреляции остатков.

Если $D_2(\alpha) > D > D_1(\alpha)$ или $4 - D_1(\alpha) > D > 4 - D_2(\alpha)$, то критерий не позволяет принять решение по гипотезе о наличии или отсутствии корреляции остатков. Если $D_2(\alpha) < D < 4 - D_2(\alpha)$, то гипотеза корреляции остатков отклоняется. Критические значения $D_1(\alpha)$ и $D_2(\alpha)$ для различных α и числа k коэффициентов в регрессии (имеется в виду число коэффициентов в регрессии $\hat{y} = a + b_1x + b_2x + \ldots + b_kx$ — такого вида множественную регрессию мы рассмотрим позже), приведены в табл. 231.

Таблица 231

Критические значения статистики Дарбина–Ватсона
(α — доверительная вероятность, k — число коэффициентов в модели $y = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k$) [591]

n	α	k = 1		2		3		4		5	
		D_1	D_2	D_1	D_2	D_1	D_2	D_1	D_2	D_1	D_2
15	0,95	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
	0,99	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,96
20	0,95	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
	0,99	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,60	1,74
25	0,95	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
	0,99	1,05	1,21	0,98	1,30	0,90	1,41	0,83	1,52	0,75	1,65
30	0,95	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
	0,99	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
40	0,95	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
	0,99	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,05	1,58
50	0,95	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
	0,99	1,32	1,40	1,28	1,45	1,24	1,49	1,20	1,54	1,16	1,59
60	0,95	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
	0,99	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60
80	0,95	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
	0,99	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62
100	0,95	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78
	0,99	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65

Задача 364. *Для полученных в результате эксперимента данных* ($n = 15$)

x_i:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15;
y_i:	7	8	6	9	11	10	14	13	18	19	11	14	18	16	16

проверить наличие корреляции регрессионных остатков критерием Дарбина–Ватсона при доверительной вероятности $\alpha = 0{,}95$.

Вычислим оценки регрессии методом наименьших квадратов (см. раздел 5.3.1.1)

$$b = \frac{n \cdot \sum x_i \cdot y_i - \sum x_i \cdot \sum y_i}{n \cdot \sum x_i^2 - \left(\sum x_i\right)^2}; \quad a = \frac{\sum y_i - b \cdot \sum x_1}{n}.$$

Тогда

$$\sum_{i=1}^{15} x_i = 120; \quad \sum_{i=1}^{15} x_i^2 = 1240; \quad \left(\sum_{i=1}^{15} x_i\right)^2 = 14400; \quad \sum_{i=1}^{15} x_i \cdot y_i = 1733;$$

$$b = \frac{15 \cdot 1733 - 120 \cdot 190}{15 \cdot 1240 - 14400} = 0{,}7607; \quad a = \frac{190 - 0{,}7607 \cdot 120}{15} = 6{,}581.$$

Находим регрессионные остатки $e_i = \hat{y}_i - y_i = 6{,}581 + 0{,}7607 \cdot x_i - y_i$:

$$0{,}3417;\ 0{,}1024;\ 2{,}8631;\ 0{,}6238;\ -0{,}6155;\ 1{,}1452;\ -2{,}0941;\ -0{,}3334;$$
$$-4{,}5727;\ -4{,}8120;\ 3{,}848;\ 1{,}7094; -1{,}5299;\ 1{,}2308;\ 1{,}9915.$$

Вычисляем статистику Дарбина–Ватсона

$$\sum_{i=2}^{15} (e_i - e_{i-1})^2 = (0{,}1024 - 0{,}3417)^2 + (2{,}8631 - 0{,}1024)^2 + \ldots$$
$$\ldots + (1{,}2308 + 1{,}5299)^2 + (1{,}9915 - 1{,}2308)^2 = 149{,}4087;$$

$$\sum_{i=1}^{15} e_i^2 = 856{,}3012; \quad D = \frac{149{,}4087}{85{,}3012} = 1{,}7515.$$

Из табл. 231 для $\alpha = 0{,}95$, $k = 1$ (так как регрессия $y = a + b \cdot x$ имеет один коэффициент регрессии b, не считая свободного члена a) и $n = 15$ имеем $D_1(0{,}95) = 1{,}08$ и $D_2(0{,}95) = 1{,}36$.

В нашем случае $D_2(0{,}95) = 1{,}36 < D = 1{,}7515 < 4 - D_1(0{,}95) = 4 - 1{,}08 = 2{,}92$.

Следовательно, наличие корреляции остатков регрессионной модели

$$y = 6,581 + 0,7606 \cdot x$$

с достоверностью $\alpha = 0{,}95$ отклоняется.

5.3.1.2.2.3. Оценка выбросов в регрессии

Наличие грубых отклонений (промахов, выбросов) в значениях y_i, не связанных с естественным разбросом, может приводить к большим ошибкам при построении регрессии. Учитывая, что в практике регрессионная модель часто используется для предсказания поведения исследуемой случайной величины, то наличие выброса в данных может привести к грубым ошибкам прогноза.

Ниже рассмотрены некоторые методы выявления выбросов в регрессии. Будем использовать следующие обозначения: $e_i = y_i - \hat{y}_i$ — остатки в точке $x = x_i$, $\bar{e} = \frac{1}{n} \sum_{i=1}^{n} e_i$.

5.3.1.2.2.3.1. Критерий Эктона

Статистикой критерия является величина [592] $V = \dfrac{|e_k - \bar{e}|}{S_k}$, где e_k — остаток от предполагаемого выброса; $\bar{e} = \dfrac{1}{n-1}\sum\limits_{\substack{i=1\\i\neq k}}^{n} e_i$ — среднее по всем другим остаткам; S_k^2 — дисперсия отклонения экспериментальных точек линии регрессии с учетом отбрасывания подозрительного (k-го) наблюдения.

Критерий Эктона применим при $n \geqslant 30$ для выделения только одного выброса в простейшей линейной модели $y = a + bx$.

Предположим, что при каждом значении независимой переменной x_i получено m_i значений зависимой переменной y_i. Тогда оценка для $S_{y_i}^2$ в нашем случае имеет вид

$$S_k^2 = \frac{\sum\limits_{\substack{i=1\\i\neq k}}^{n}\sum\limits_{j=1}^{m_i}(y_{ij} - \bar{y}_i)^2 + \sum\limits_{\substack{i=1\\i\neq k}}^{n} m_i(\bar{y}_i - \hat{y}_i)^2}{\sum\limits_{\substack{i=1\\i\neq k}}^{n} m_i - 2}.$$

Остаток e_i с вероятностью α признается выбросом, если $V > V_\alpha$, где V_α — критическое значение, приведенное в табл. 232.

Таблица 232

Критические значения V_α критерия выбросов Эктона

(α — доверительная вероятность) [14]

n	α		n	α		n	α	
	0,95	0,99		0,95	0,99		0,95	0,99
3	123	31,4	7	3,98	5,88	15	3,34	4,22
4	7,17	16,27	8	3,77	5,33	20	3,28	4,02
5	5,05	9	9	3,63	4,98	25	3,26	3,94
6	4,34	6,85	10	3,54	4,75			

Задача 365. *В результате эксперимента были получены следующие значения y_i, соответствующие различным значениям независимой переменной x_i:*

x_i:	1,2	2,7	3,9	5,1	7,2	9,4	11,1	12,4	16,1	19,1
y_i:	3,1	6,1	12,1	21,3	27,4	31,2	41,1	51,2	71,1	91,2
	4,1	8,4	16,1	19,4	21,5	36,4	38,2	48,1	59,2	81,2
	5,6	9,2	9,3	18,1	19,8	27,3	31,4	31,4	63,5	79,4
$\bar{y}_i$	4,27	7,9	12,5	19,6	22,9	31,63	36,9	43,56	64,6	83,93.

Необходимо проверить наличие выброса в регрессионной модели при доверительной вероятности $\alpha = 0{,}95$ критерием Эктона.

Вычислим параметры регрессионной модели (см. раздел 5.3.1.1.1):

$$\sum_{i=1}^{10} x_i = 88{,}2; \quad \left(\sum_{i=1}^{10} x_i\right)^2 = 7779{,}24; \quad \sum_{i=1}^{10} x_i^2 = 1091{,}14;$$

$$\sum_{i=1}^{10} \bar{y}_i = 327{,}79; \quad \sum_{i=1}^{10} x_i \cdot \bar{y}_i = 4230{,}223;$$

$$b = \frac{10 \cdot 4230{,}223 - 88{,}2 \cdot 327{,}79}{10 \cdot 1091{,}14 - 7779{,}24} = 4{,}275; \quad a = \frac{327{,}79 - 4{,}275 \cdot 88{,}2}{10} = -4{,}9265.$$

Остатки равны $e_i = \bar{y}_i - \hat{y}_i \cdot (\hat{y}_i = -4,9265 + 4,275 \cdot x_i)$:

4,066; 1,284; 0,754; 2,724; −2,953; −3,628; −5,626; −4,523; 0,699; 7,204.

Проверяем на выброс максимальный по модулю остаток $e_{10} = 7{,}204$. Исключая соответствующую ему точку $x_{10} = 19{,}1$, вычисляем $S_{10}^2(m_i = 3)$:

$$\sum_{i=1}^{9} \sum_{j=1}^{3} (y_{ij} - \bar{y}_i)^2 = (3{,}1 - 4{,}27)^2 + (4{,}1 - 4{,}27)^2 + \ldots + (63{,}5 - 64{,}6)^2 = 459{,}2302;$$

$$3 \cdot \sum_{i=1}^{9} (\bar{y}_i - \hat{y}_i)^2 = 301{,}9509; \quad S_{10}^2 = \frac{459{,}2302 + 301{,}9509}{25} = 30{,}477 \ (S_{10} = 5{,}518).$$

Среднее по $(n-1)$ остаткам, не считая потенциального выброса, равно

$$\frac{1}{9} \cdot \sum_{i=1}^{9} e_i = -0{,}80.$$

Вычисляем статистику критерия Эктона $V = \dfrac{7{,}204 + 0{,}80}{5{,}518} = 1{,}45$.

Из табл. 232 для $n = 10$ и $\alpha = 0{,}95$ имеем $V_{0,95} = 3{,}54$.

Так как $V = 1{,}45 < V_{0,95} = 3{,}54$, гипотезу о наличии выбросов в регрессии следует отклонить.

5.3.1.2.2.3.2. Критерий Титьена–Мура–Бекмана

В работе Титьена–Мура–Бекмана [593] предложен критерий обнаружения одного выброса в линейной модели $y = a + bx$, основанный на статистике

$$R = \max\left|\frac{e_i}{S_i}\right|, \quad \text{где} \quad S_i^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right].$$

Если $R > R_\alpha$, то y_i, соответствующее максимальному значению отношения $\dfrac{e_i}{S_i}$, признается с вероятностью α выбросом. Критические значения R_α приведены в табл. 233.

Таблица 233

Критические значения R_α критерия выбросов в регрессии Титьена–Мура–Бекмана [593]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
4	1,41	1,41	1,41	16	2,50	2,64	2,92
5	1,69	1,71	1,73	18	2,56	2,71	2,99
6	1,88	1,92	1,97	20	2,60	2,76	3,06
7	2,01	2,07	2,16	24	2,69	2,85	3,17
8	2,10	2,19	2,31	30	2,79	2,97	3,28
9	2,18	2,28	2,43	36	2,86	3,03	3,35
10	2,24	2,35	2,53	48	2,97	3,15	3,41
11	2,30	2,43	2,64	60	3,04	3,21	3,50
12	2,35	2,48	2,70	100	3,22	3,40	3,75
14	2,43	2,57	2,80				

Задача 366. *Для данных задачи 365 проверить наличие выброса в регрессионной модели критерием Титьена–Мура–Бекмана при $\alpha = 0{,}95$.*

Имеем (используем значения e_i, вычисленные ранее в задаче 365)

$$S_i^2 = \frac{\sum\limits_{i=1}^{10} e_i^2}{8} \cdot \left[1 - \frac{1}{10} \cdot \frac{(x_i - \bar{x})^2}{\sum\limits_{i=1}^{10} (x_i - \bar{x})^2}\right] =$$

$$= \frac{152{,}5479}{8} \cdot \left[1 - 0{,}1 \cdot \frac{(x_i - \bar{x})^2}{313{,}216}\right] = 17{,}1616 - 0{,}06088 \cdot (x_i - 8{,}82)^2\,.$$

Вычисляем последовательность значений $\dfrac{e_i}{S_i}$:

$$\frac{e_1}{S_1} = \frac{4{,}066}{\sqrt{17{,}1616 - 0{,}06088 \cdot (1{,}28{,}82)}} = 1{,}01 \quad \text{и по аналогии}$$

$$\frac{e_2}{S_2} = 0{,}333; \quad \frac{e_3}{S_3} = 0{,}190; \quad \frac{e_4}{S_4} = 0{,}674; \quad \frac{e_5}{S_5} = -0{,}716; \quad \frac{e_6}{S_6} = -0{,}876;$$

$$\frac{e_7}{S_7} = -1{,}371; \quad \frac{e_8}{S_8} = -1{,}117; \quad \frac{e_9}{S_9} = 0{,}187; \quad \frac{e_{10}}{S_{10}} = 2{,}199.$$

Находим $R = \max\left|\dfrac{e_i}{S_i}\right| = 2{,}199$. Из табл. 233 для $\alpha = 0{,}95$ и $n = 2{,}35$ находим $R_{0,95} = 2{,}35$. Так как $R = 2{,}199 < R_{0,95} = 2{,}35$, гипотеза о наличии выброса отклоняется.

5.3.1.2.2.3.3. Критерий Прескотта–Лунда

Прескотт [594] и Лунд [595], учитывая близость значений S_i (собственно, эта близость является одной из предпосылок регрессионного анализа — см. вводную часть раздела 5.3), предложили упрощенную модификацию критерия Титьена–Мура–Бекмана, основанную на статистике

$$R^* = \max\left|\frac{e_i}{\bar{S}}\right|,$$

где $\bar{S}^2$ — оценка средней дисперсии остатков $\left(\bar{S}^2 = \dfrac{n-q}{n} S^2\right)$; q — количество параметров в регрессионной модели.

Более удобна эквивалентная форма критерия Прескотта–Лунда

$$R^* = \sqrt{n}\max\frac{|e_i|}{\sqrt{\sum\limits_{i=1}^{n} e_i^2}}.$$

Гипотеза наличия выброса в линейной регрессионной модели с k параметрами (ранее рассматривался случай простейшей модели, когда $k = 2$; случай $k > 2$ параметров — множественная регрессия — будет рассмотрен далее в разделе 5.3.2) отклоняется, если $R^* < R^*_\alpha$, где R^*_α — критическое значение, равное

$$R^*_\alpha = \sqrt{\frac{(n-k)F}{n-k+1+F}};$$

F — $\left(1 - \dfrac{1-\alpha}{n}\right)$-квантиль F-распределения Фишера с $f_1 = 1$ и $f_2 = n - k - 1$ степенями свободы.

Таблицы критических значений R^*_α приведены Лундом [595] и воспроизведены в табл. 234.

Таблица 234

Критические значения R^*_α критерия выбросов в регрессии Прескотта–Лунда [595]

n	k							
	1	2	3	4	5	6	8	10
Доверительная вероятность $\alpha = 0{,}90$								
5	1,87							
6	2,00	1,89						
7	2,10	2,02	1,90					
8	2,18	2,12	2,03	1,91				
9	2,24	2,20	2,13	2,05	1,92			
10	2,30	2,26	2,21	2,15	2,06	1,92		
12	2,39	2,37	2,33	2,29	2,24	2,17	1,93	
14	2,47	2,45	2,42	2,39	2,36	2,31	2,19	1,94
16	2,53	2,51	2,50	2,47	2,45	2,42	2,34	2,23
18	2,58	2,57	2,56	2,54	2,52	2,50	2,44	2,35
20	2,63	2,62	2,61	2,59	2,58	2,56	2,52	2,46
25	2,72	2,72	2,71	2,70	2,69	2,68	2,66	2,63
30	2,80	2,79	2,77	2,78	2,77	2,77	2,75	2,73
35	2,86	2,85	2,85	2,85	2,84	2,84	2,82	2,81
40	2,91	2,91	2,90	2,90	2,90	2,89	2,88	2,87
45	2,95	2,95	2,95	2,95	2,94	2,94	2,93	2,93
50	2,99	2,99	2,99	2,99	2,98	2,98	2,98	2,97
60	3,06	3,06	3,05	3,05	3,05	3,05	3,05	3,04
70	3,11	3,11	3,11	3,11	3,11	3,11	3,10	3,10
80	3,16	3,16	3,15	3,15	3,15	3,15	3,15	3,15
90	3,20	3,20	3,19	3,19	3,19	3,19	3,19	3,19
100	3,23	3,23	3,23	3,23	3,23	3,23	3,23	3,22
Доверительная вероятность $\alpha = 0{,}95$								
5	1,92							
6	2,07	1,93						
7	2,19	2,08	1,94					
8	2,28	2,20	2,10	1,94				
9	2,35	2,29	2,21	2,10	1,95			
10	2,42	2,37	2,31	2,22	2,11	1,95		
12	2,52	2,49	2,45	2,39	2,33	2,24	1,96	
14	2,61	2,58	2,55	2,51	2,47	2,41	2,25	1,96
16	2,68	2,66	2,63	2,60	2,57	2,53	2,43	2,26
18	2,73	2,72	2,70	2,68	2,65	2,62	2,55	2,44
20	2,78	2,77	2,76	2,74	2,72	2,70	2,64	2,57
25	2,89	2,88	2,87	2,86	2,84	2,83	2,80	2,76
30	2,96	2,96	2,95	2,94	2,93	2,93	2,90	2,88
35	3,03	3,02	3,02	3,01	3,00	3,00	2,98	2,97
40	3,08	3,08	3,07	3,07	3,06	3,06	3,05	3,03
45	3,13	3,12	3,12	3,12	3,11	3,11	3,10	3,09
50	3,17	3,16	3,16	3,16	3,15	3,15	3,14	3,14
60	3,23	3,23	3,23	3,23	3,22	3,22	3,22	3,21
70	3,29	3,29	3,28	3,28	3,28	3,28	3,27	3,27
80	3,33	3,33	3,33	3,33	3,33	3,33	3,32	3,32
90	3,37	3,37	3,37	3,36	3,37	3,38	3,36	3,36
100	3,41	3,41	3,40	3,40	3,40	3,40	3,40	3,40

Задача 367. *Для данных задачи 365 проверить наличие выброса в регрессионной модели критерием Прескотта–Лунда при доверительной вероятности* $\alpha = 0{,}90$.

Пользуясь оценками и расчетами, выполненными в задаче 365, получим ряд значений $\frac{|e_i|}{\sqrt{\sum\limits_{i=1}^{n} e_i^2}}$ $(i = 1, 2, \ldots, n) \to$

0,3292; 0,2563; 0,0610; 0,2200; 0,2390; 0,2937; 0,4555; 0,3662; 0,0566; 0,5830.

Имеем

$$R^* = \sqrt{10} \cdot \max \frac{|e_i|}{\sqrt{\sum\limits_{i=1}^{n} e_i^2}} = \sqrt{10} \cdot 0{,}583 = 1{,}844.$$

Из табл. 234 для $\alpha = 0{,}90$, $k = 2$ и $n = 10$ имеем $R^*_{0,90} = 2{,}26$.

Так как $R^* = 1{,}844 < R^*_{0,90} = 2{,}26$, гипотеза о наличии выброса в регрессии отклоняется.

Воспользуемся квантилями F-распределения.

Находим по таблицам F-распределения (при необходимости интерполяцией)

$$F_{1-\frac{1-\alpha}{n}} = F_{1-\frac{1-0,9}{10}} = F_{0,99}(f_1 = 1; f_2 = n - 2 - 1) = F_{0,99}(1; 7) = 12{,}246.$$

Тогда $R^*_{0,90} = \sqrt{\frac{(n-k)\cdot F}{n-k-1+F}} = \sqrt{\frac{(10-2)\cdot 12{,}246}{10-2-1+12{,}246}} = 2{,}256$, что достаточно близко к табличному значению $R^*_{0,90} = 2{,}26$.

5.3.1.2.3. Доверительные области и толерантные границы регрессии

5.3.1.2.3.1. Доверительная область простой линейной регрессии

Для линейной регрессии $y = a + bx$ доверительная область (с доверительной вероятностью α) на заданном отрезке будет ограничена гиперболами

$$a + bx \pm u(\nu, \beta, \lambda)\, S \sqrt{\frac{(x-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2} + \frac{1}{n}},$$

где

$$\nu = n - 2; \quad \lambda = \left\{\frac{1}{2}\left[1 - \frac{1 + nCD}{\left(1 + nC^2\right)\left(1 + nD^2\right)}\right]\right\}^{\frac{1}{2}};$$

$$C = \frac{x_1 - \bar{x}}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}}; \quad D = \frac{x_2 - \bar{x}}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}}; \quad S_y^2 = \frac{1}{\nu}\sum_{i=1}^{n}(y_i - a - bx_i)^2;$$

$u(\nu, \alpha, \lambda)$ — коэффициент, зависящий от ν, α и λ.

Значения $u(\nu, \alpha, \lambda)$ приведены в табл. 235.

Доверительная зона, рассчитанная по приведенным выше соотношениям, является геометрическим местом доверительных интервалов для различных значений x.

Коэффициенты $u(\nu, \alpha, \lambda)$ для построения

ν	λ									
	0	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45
					$\alpha = 0{,}90$					
1	6,314	6,625	6,922	7,203	7,470	7,724	7,965	8,192	8,407	8,609
2	2,920	3,022	3,121	3,216	3,308	3,397	3,482	3,564	3,642	3,717
3	2,353	2,425	2,494	2,561	2,625	2,687	2,747	2,805	2,860	2,914
4	2,132	2,192	2,250	2,306	2,360	2,412	2,463	2,511	2,558	2,604
5	2,015	2,071	2,123	2,173	2,222	2,268	2,313	2,357	2,399	2,440
6	1,943	1,995	2,044	2,091	2,136	2,180	2,222	2,263	2,302	2,340
7	1,895	1,944	1,991	2,036	2,079	2,120	2,160	2,199	2,236	2,272
8	1,860	1,908	1,953	1,997	2,038	2,078	2,116	2,153	2,189	2,224
9	1,833	1,880	1,925	1,967	2,007	2,046	2,083	2,119	2,153	2,287
10	1,812	1,859	1,902	1,944	1,983	2,021	2,057	2,092	2,126	2,159
11	1,796	1,841	1,884	1,925	1,963	2,001	2,036	2,071	2,104	2,136
12	1,782	1,827	1,869	1,909	1,947	1,984	2,019	2,053	2,086	2,117
13	1,771	1,815	1,857	1,897	1,934	1,970	2,005	2,038	2,070	2,102
14	1,761	1,805	1,847	1,886	1,923	1,959	1,993	2,026	2,058	2,089
15	1,753	1,797	1,838	1,876	1,913	1,948	1,982	2,015	2,046	2,077
16	1,746	1,789	1,830	1,868	1,905	1,940	1,973	2,006	2,037	2,067
17	1,740	1,783	1,823	1,861	1,897	1,932	1,965	1,998	2,029	2,059
18	1,734	1,777	1,817	1,855	1,891	1,925	1,959	1,990	2,021	2,051
19	1,729	1,772	1,811	1,849	1,885	1,919	1,952	1,984	2,015	2,044
20	1,725	1,767	1,807	1,844	1,880	1,914	1,947	1,978	2,009	2,038
25	1,708	1,750	1,789	1,825	1,860	1,894	1,926	1,957	1,986	2,015
50	1,676	1,716	1,753	1,789	1,823	1,855	1,886	1,915	1,943	1,971
100	1,660	1,700	1,736	1,771	1,804	1,836	1,866	1,895	1,923	1,950
∞	1,645	1,684	1,720	1,754	1,786	1,817	1,847	1,875	1,902	1,929
					$\alpha = 0{,}95$					
1	12,700	13,320	13,920	14,480	15,010	15,520	16,000	16,460	16,890	17,290
2	4,303	4,445	4,582	4,715	4,844	4,968	5,087	5,202	5,312	5,417
3	3,182	3,271	3,356	3,438	3,518	3,594	3,668	3,740	3,809	3,875
4	2,776	2,847	2,915	2,978	3,040	3,101	3,159	3,215	3,270	3,332
5	2,571	2,633	2,692	2,748	2,802	2,854	2,904	2,953	3,000	3,045
6	2,447	2,504	2,558	2,609	2,659	2,706	2,751	2,795	2,838	2,879
7	2,365	2,418	2,469	2,517	2,563	2,607	2,650	2,691	2,730	2,769
8	2,306	2,357	2,406	2,452	2,496	2,538	2,578	2,616	2,654	2,690
9	2,262	2,312	2,358	2,403	2,445	2,485	2,524	2,561	2,597	2,632
10	2,228	2,277	2,322	2,365	2,406	2,445	2,482	2,518	2,552	2,586
11	2,201	2,249	2,293	2,335	2,375	2,413	2,449	2,484	2,517	2,500
12	2,179	2,226	2,269	2,311	2,350	2,387	2,422	2,456	2,489	2,521
13	2,160	2,207	2,250	2,290	2,329	2,365	2,400	2,433	2,466	2,497
14	2,145	2,191	2,233	2,273	2,331	2,347	2,381	2,414	2,446	2,476
15	2,132	2,177	2,219	2,258	2,295	2,331	2,365	2,397	2,429	2,459
16	2,120	2,165	2,206	2,245	2,282	2,317	2,351	2,383	2,414	2,444
17	2,110	2,154	2,195	2,234	2,271	2,305	2,338	2,370	2,401	2,431
18	2,101	2,145	2,186	2,224	2,260	2,295	2,328	2,359	2,389	2,419
19	2,093	2,137	2,177	2,215	2,251	2,285	2,318	2,349	2,379	2,408
20	2,086	2,129	2,170	2,207	2,243	2,277	2,309	2,340	2,370	2,399
25	2,060	2,102	2,141	2,178	2,213	2,246	2,277	2,307	2,336	2,364
50	2,009	2,049	2,086	2,121	2,154	2,186	2,215	2,244	2,271	2,297
100	1,984	2,023	2,060	2,094	2,126	2,157	2,186	2,213	2,240	0,265
∞	1,960	1,998	2,034	2,068	2,099	2,128	2,156	2,183	2,209	2,234

Таблица 235

доверительной области линейной регрессии [25]

λ										
0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	1,00
$\alpha = 0{,}90$										
8,799	8,976	9,142	9,294	9,434	9,561	9,674	9,772	9,854	9,916	9,950
3,788	3,855	3,919	3,978	4,033	4,083	4,129	4,169	4,202	4,228	4,243
2,965	3,015	3,061	3,105	3,146	3,184	3,218	3,248	3,274	3,294	3,305
2,647	2,689	2,728	2,766	2,801	2,834	2,864	2,891	2,913	2,931	2,941
2,480	2,517	2,554	2,588	2,620	2,650	2,678	2,703	2,724	2,740	2,749
2,377	2,412	2,446	2,478	2,509	2,538	2,564	2,587	2,607	2,623	2,632
2,308	2,341	2,374	2,405	2,434	2,462	2,487	2,509	2,529	2,545	2,552
2,258	2,290	2,322	2,352	2,381	2,407	2,432	2,453	2,472	2,487	2,495
2,220	2,252	2,283	2,312	2,340	2,366	2,390	2,411	2,430	2,444	2,452
2,191	2,222	2,252	2,281	2,308	2,333	2,357	2,378	2,397	2,411	2,418
2,168	2,198	2,227	2,256	2,283	2,308	2,331	2,352	2,370	2,384	2,392
2,148	2,178	2,207	2,235	2,262	2,287	2,310	2,330	2,348	2,362	2,269
2,132	2,162	2,190	2,218	2,244	2,269	2,292	2,312	2,330	2,343	2,351
2,119	2,148	2,176	2,203	2,229	2,254	2,276	2,297	2,314	2,328	2,335
2,107	2,136	2,164	2,191	2,217	2,241	2,263	2,283	2,301	2,314	2,322
2,088	2,125	2,154	2,191	2,217	2,241	2,263	2,283	2,301	2,314	2,322
2,080	2,116	2,144	2,171	2,196	2,220	2,242	2,262	2,279	2,293	2,300
2,073	2,108	2,136	2,162	2,187	2,211	2,233	2,253	2,270	2,284	2,291
2,067	2,101	2,128	2,155	2,180	2,203	2,225	2,245	2,262	2,276	2,283
2,044	2,095	2,122	2,148	2,173	2,196	2,218	2,238	2,255	2,268	2,276
1,998	2,071	2,097	2,123	2,147	2,170	2,192	2,211	2,228	2,241	2,249
1,976	2,024	0,050	2,074	2,098	2,120	2,141	2,160	2,176	2,189	2,196
1,954	2,002	2,026	2,050	2,074	2,096	2,116	2,135	2,151	2,164	2,171
	1,979	2,004	2,027	2,050	2,072	2,092	2,110	2,126	2,139	2,146
$\alpha = 0{,}95$										
17,670	10,025	18,355	18,661	18,941	19,195	19,422	19,619	19,782	19,907	19,975
5,517	5,612	5,701	5,785	5,864	5,935	6,000	6,058	6,107	6,144	6,164
3,939	4,000	4,059	4,114	4,166	4,214	4,258	4,297	4,330	4,356	4,371
3,373	3,422	3,469	3,514	3,556	3,596	3,632	3,664	3,692	3,714	3,727
3,089	3,132	3,173	3,212	3,249	3,284	3,316	3,345	3,371	3,390	3,402
2,919	2,958	2,995	3,031	3,065	3,097	3,127	3,154	3,178	3,197	3,207
2,806	2,842	2,877	2,911	2,943	2,974	3,002	3,028	3,051	3,068	3,078
2,726	2,760	2,793	2,826	2,856	2,886	2,913	2,938	2,959	2,976	2,986
2,666	2,699	2,731	2,762	2,792	2,820	2,846	2,870	2,891	2,908	2,918
2,618	2,650	2,681	2,712	2,741	2,768	2,794	2,818	2,838	2,855	2,865
2,582	2,613	2,643	2,673	2,701	2,728	2,753	2,777	2,797	2,813	2,823
2,552	2,582	2,612	2,641	2,669	2,695	2,720	2,743	2,763	2,779	2,788
2,527	2,557	2,586	2,614	2,642	2,668	2,692	2,715	2,734	2,750	2,759
2,506	2,536	2,564	2,592	2,619	2,645	2,669	2,691	2,710	2,726	2,735
2,488	2,517	2,545	2,573	2,599	2,625	2,649	2,671	2,690	2,706	2,714
2,473	2,501	2,529	2,556	2,583	2,608	2,631	2,653	2,672	2,688	2,696
2,459	2,487	2,515	2,542	2,568	2,593	2,616	2,638	2,657	2,672	2,680
2,447	2,475	2,502	2,529	2,555	2,579	2,603	2,624	2,643	2,658	2,666
2,437	2,464	2,491	2,518	2,543	2,568	2,591	2,612	2,631	2,646	2,654
2,427	2,455	2,481	2,508	2,533	2,577	2,580	2,601	2,620	2,635	2,643
2,392	2,418	2,444	2,470	2,494	2,518	2,540	2,561	2,579	2,594	2,602
2,323	2,348	2,372	2,396	2,419	2,442	2,463	2,483	2,501	2,515	2,523
2,290	2,314	2,338	2,361	2,384	2,405	2,426	2,446	2,463	2,248	2,485
2,258	2,281	2,304	2,327	2,349	2,370	2,390	2,409	2,426	2,440	2,448

Для единичного значения $\hat{y}_i$ доверительный интервал будет равен

$$\hat{y}_i \pm t_{\frac{1+\alpha}{2}} \pm S_i,$$

где

$$S_i = S_y \left[1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\right]^{\frac{1}{2}}; \quad S_y = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}; \quad \hat{y}_i = a + bx_i.$$

Легко видеть, что этот интервал будет намного шире, чем границы доверительной зоны для всей регрессии.

Задача 368. *По данным задачи 365 построить 95%-ю доверительную зону для регрессии на интервале от $x_1 = 3$ до $x_2 = 10$ и доверительный интервал для $\hat{y}_i$ при $x_i = 5{,}1$ ($\bar{x} = 8{,}82$).*

Имеем

$$\nu = 10 - 2 = 8; \quad C = \frac{3 - 8{,}82}{\sqrt{\sum\limits_{i=1}^{10}(x_i - \bar{x})^2}} = \frac{-5{,}82}{17{,}6979} = -0{,}329;$$

$$D = \frac{10 - 8{,}82}{17{,}6979} = 0{,}066; \quad S_y^2 = \frac{1}{n-2}\cdot\sum_{i=1}^{10}(y_i - \hat{y}_i)^2 = 30{,}447; \quad S_y = 5{,}518;$$

$$\lambda = \sqrt{\frac{1}{2}\cdot\left[1 - \frac{(1 + 10\cdot(-0{,}329)\cdot 0{,}0666)}{(1 + 10\cdot 0{,}329^2)\cdot(1 + 10\cdot 0{,}0666^2)}\right]} = \sqrt{\frac{1}{2}\cdot\left(1 - \frac{0{,}780886}{2{,}08241\cdot 0{,}04435}\right)} = 0{,}566.$$

Из табл. 235 для $\nu = 8$, $\alpha = 0{,}95$ и $\lambda = 0{,}566$ находим (интерполяцией) $u(8; 0{,}95; 0{,}566) = 2{,}660$.

Тогда доверительная зона для регрессии имеет вид (ранее в задаче 365 было получено уравнение $\hat{y} = -4{,}9265 + 4{,}275\cdot x$): $y_н \leqslant \hat{y} \leqslant y_в$, где

$$y_н = -4{,}9265 + 4{,}275\cdot x - 2{,}66\cdot 5{,}518\cdot\left\{\frac{(x-\bar{x})^2}{313{,}2156} + \frac{1}{10}\right\}^{\frac{1}{2}};$$

$$y_в = -4{,}9265 + 4{,}275\cdot x + 2{,}66\cdot 5{,}518\cdot\left\{\frac{(x-\bar{x})^2}{313{,}2156} + \frac{1}{10}\right\}^{\frac{1}{2}}.$$

Для значения $\hat{y}_i = y(5{,}1)$ имеем (при $f = n - 2 = 8$ степенях свободы) $t_{(1+0{,}95)/2} = t_{0{,}975} = 2{,}306$ и

$$S = S_{y_i}\cdot\left\{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\right\}^{\frac{1}{2}} = 5{,}518\cdot\left\{1 + \frac{1}{10} + \frac{(5{,}1 - 8{,}82)^2}{313{,}2156}\right\}^{\frac{1}{2}} = 5{,}902;$$

$$-0{,}9265 + 4{,}275\cdot 5{,}1 - 2{,}306\cdot 5{,}902 =$$
$$= 3{,}266 \leqslant \hat{y}\,(5{,}1) \leqslant 30{,}486 = -4{,}9265 + 4{,}275\cdot 5{,}1 + 2{,}306\cdot 5{,}902.$$

Соответствующие границы доверительной зоны будут (подставляем $x = 5{,}1$ в уравнение гиперболы):

$$\hat{y} \geqslant -4{,}9265 + 4{,}275\cdot 5{,}1 - 14{,}678\cdot\left\{\frac{(5{,}1 - 8{,}82)^2}{313{,}2156} + \frac{1}{10}\right\}^{\frac{1}{2}} = 11{,}30;$$

$$\hat{y} \leqslant -4{,}9265 + 4{,}275\cdot 5{,}1 + 14{,}678\cdot\left\{\frac{(5{,}1 - 8{,}82)^2}{313{,}2156} + \frac{1}{10}\right\}^{\frac{1}{2}} = 22{,}449.$$

Видим, что длина доверительного интервала в этом случае уменьшается в $\dfrac{30{,}486 - 3{,}266}{22{,}449 - 11{,}30} = 2{,}4$ раза.

5.3.1.2.3.2. Оценка обращенного уравнения регрессии

Иногда возникает так называемая обращенная задача оценивания, заключающаяся в ответе на вопрос: если найдена оценка линии регрессии, то каким образом по некоторому измеренному значению y можно предсказать значение переменной x_0, не являющейся случайной величиной? Браунли [559] и Хальд [51] показали, что α-доверительный интервал для оценивания значения x_0 определяется выражением

$$\bar{x} + \frac{\bar{y}_0 - \bar{y}}{b^*} - t_{\frac{1+\alpha}{2}} \frac{S}{b^*}\sqrt{p} \leqslant x_0 \leqslant \bar{x} + \frac{\bar{y}_0 - \bar{y}}{b^*} + t_{\frac{1+\alpha}{2}} \frac{S}{b^*}\sqrt{p},$$

где $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$; $\bar{y} = \dfrac{\sum\limits_{i=1}^{n} m_i \bar{y}_i}{\sum\limits_{i=1}^{n} m_i}$; $\bar{y}_i = \dfrac{1}{m_i}\sum\limits_{j=1}^{m_i} y_{ij}$; $b^* = b - \dfrac{t^2_{\frac{1+\alpha}{2}} S^2}{b\sum\limits_{i=1}^{n} m_i\,(x_i - \bar{x})^2}$;

$$p = \left(\frac{1}{m_0} + \frac{1}{\sum\limits_{i=1}^{n} m_i}\right)\frac{b^*}{b} + \frac{(x^* - \bar{x})^2}{\sum\limits_{i=1}^{n} m_i\,(x_i - \bar{x})^2}; \quad x^* = \frac{\bar{y}_0 - \bar{y}}{b}\bar{x}; \quad S^2 = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m_i} (y_{ij} - a - bx_i)^2}{\sum\limits_{i=1}^{n} m_i - 2};$$

m_i — количество наблюдений y_i при $x = x_i$; m_0 — количество наблюдений y_i при $x = x_0$; $\bar{y}_0$ — среднее по m_0 наблюдениям y_i при $x = x_0$; t_α — α-квантиль распределения Стьюдента при $f = n - 2$ степенях свободы.

Приведенный доверительный интервал основан на совместном распределении линии регрессии и $\bar{y}_0$. Он неприменим для вычисления значений x, соответствующих ряду значений y, если используется одна и та же линия регрессии.

Укажем еще один способ оценки x_0, предложенный Крючковым и Эйзенхартом [14], в соответствии с которым оценкой x_0 служит величина $\widehat{x}_0 = c + d\bar{y}_0$, где $d = \dfrac{\sum\limits_{i=1}^{n} m_i(x_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum\limits_{i=1}^{n} m_i(\bar{y}_i - \tilde{y})^2}$; $c = \bar{x} - d\bar{y}$.

Задача 369. *По данным задачи 365 найти 95%-й доверительный интервал для величины x_0, если при $x = x_0$ были получены следующие значения y_{0i}: 28,1; 29,2; 25,6.*

Имеем

$$\bar{x} = 8{,}82; \quad m_i = 3; \quad \bar{y} = \frac{\sum\limits_{i=1}^{10} 3\cdot\bar{y}_i}{\sum\limits_{i=1}^{10} m_i} = \frac{1}{10}\cdot\sum_{i=1}^{10}\bar{y}_i = 32{,}779; \quad S^2 = 30{,}447\ (S = 0{,}518);$$

$$m_0 = 3; \quad b - 4{,}275; \quad t_{0,975}(8) = 2{,}306; \quad \sum_{i=1}^{10} m_i\cdot(x_i - \bar{x})^2 = 939{,}638;$$

$$\bar{y}_0 = \frac{28{,}1 + 29{,}2 + 25{,}6}{3} = 27{,}633.$$

Далее

$$x^* = \frac{\bar{y}_0 - \bar{y}}{b} + \bar{x} = \frac{27{,}633 - 32{,}779}{4{,}275} + 8{,}82 = 7{,}616; \quad b^* = \frac{4{,}275 - 2{,}306^2 \cdot 30{,}447}{4{,}275 \cdot 939{,}638} = 4{,}2347;$$

$$p = \left(\frac{1}{3} + \frac{1}{30}\right) \cdot \frac{4{,}2347}{4{,}275} + \frac{(7{,}616 - 8{,}82)^2}{939{,}638} = 0{,}36475.$$

Вычисляем:

$$\bar{x} + \frac{\bar{y}_0 - \bar{y}}{b^*} + t_{\frac{1+\alpha}{2}} \cdot \frac{S}{b^*} \cdot \sqrt{p} = 8{,}82 + \frac{27{,}633 - 32{,}779}{4{,}2347} + \frac{2{,}306 \cdot 5{,}518 \cdot \sqrt{0{,}36475}}{4{,}2347} = 9{,}419;$$

$$\bar{x} + \frac{\bar{y}_0 - \bar{y}}{b^*} - t_{\frac{1+\alpha}{2}} \cdot \frac{S}{b^*} \cdot \sqrt{p} = 8{,}82 + \frac{27{,}633 - 32{,}779}{4{,}2347} - \frac{2{,}306 \cdot 5{,}518 \cdot \sqrt{0{,}36475}}{4{,}2347} = 5{,}990.$$

Следовательно, 95%-й доверительный интервал для значения x_0, соответствующий измеренным значениям y_{0i}, имеет вид $5{,}790 \leqslant x_0 \leqslant 9{,}419$.

Для нахождения оценки Крючкова–Эйзенхарта вычислим

$$\sum_{i=1}^{n} m_i \cdot (x_i - \bar{x}) \cdot (y_i - \bar{y}) = 3 \cdot [(1{,}2 - 8{,}82) \cdot (4{,}27 - 32{,}779) + \ldots$$
$$\ldots + (19{,}1 - 8{,}82) \cdot (83{,}93 - 32{,}779)] = 4017{,}3456;$$

$$\sum_{i=1}^{10} m_i \cdot (\bar{y}_i - \bar{y})^2 = 3 \cdot \left[(4{,}27 - 32{,}779)^2 + \ldots + (83{,}93 - 32{,}779)^2\right] = 17633{,}3396.$$

Тогда имеем

$$d = \frac{4017{,}3456}{17633{,}3396} = 0{,}2278; \quad c = \bar{x} - d \cdot \bar{y} = 8{,}82 + 0{,}2278 \cdot 32{,}779 = 1{,}3529.$$

Следовательно, $\hat{x}_0 = 1{,}3529 + 0{,}2278 \cdot 27{,}633 = 7{,}65$, что очень близко к середине доверительного интервала Хальда–Брауна (см. предыдущую задачу, где $\hat{x} = \frac{9{,}419 + 5{,}790}{2} = 7{,}6045$).

5.3.1.2.3.3. Толерантные интервалы для линейной регрессии

Ранее (см. раздел 4.4) были подробно рассмотрены методы вычисления толерантных пределов для совокупности случайных величин. Напомним, что $\gamma \cdot 100\%$-м толерантным интервалом называется интервал, в границах которого находится $\gamma \cdot 100\%$ всех возможных значений случайной величины.

В настоящем разделе рассматривается методика построения толерантных интервалов для линейной регрессии.

Для того, чтобы вычислить толерантные пределы для линейной регрессии, необходимо знать дисперсию величины для разных значений независимой переменной x [596].

Эта дисперсия задается формулой Воркинга–Хотеллинга [597]

$$\sigma^2_{\hat{y}(x)} = \sigma^2 \left\{ \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \right\}.$$

Здесь эффективное число наблюдений для некоторой величины x равно

$$n^*(x) = \frac{n \sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2 + n\,(x - \bar{x})^2}.$$

Другими словами, для некоторой величины x средняя величина y определяется расчетом по линии регрессии столь же точно, как если бы было сделано n^* наблюдений при заданной величине x.

Теперь для того, чтобы найти толерантный интервал, о котором мы могли бы с вероятностью α утверждать, что внутри него лежит по крайней мере β-часть популяции, необходимо найти толерантные интервалы для каждого значения x. Искомая толерантная область будет геометрическим местом точек толерантных интервалов для отдельных x_i.

Для определения частных толерантных интервалов следует воспользоваться методами, изложенными в разделе 4.4.1.2.3 для случая выборок из нормального распределения с обоими неизвестными параметрами. Односторонние толерантные интервалы равны (см. раздел 4.4.1.2.3):

$$\text{верхний:}\quad \hat{y}_{\text{в}} = \hat{y}(x) + S(x)k(n^*(x), \alpha, \beta);$$
$$\text{нижний:}\quad \hat{y}_{\text{н}} = \hat{y}(x) - S(x)k(n^*(x), \alpha, \beta).$$

Двусторонний толерантный интервал имеет вид:

$$\hat{y}(x) - k(n^*(x), \alpha, \beta)S(x) \leqslant \hat{y}(x) \leqslant \hat{y}(x) + k(n^*(x), \alpha, \beta)S(x).$$

Эффективный объем выборки $n^*(x)$ рассчитывается по вышеприведенной формуле, исходя из значений n и $\sum\limits_{i=1}^{n}(x_i - \tilde{x})^2$.

Значение $S(x)$ рассчитывается по формуле

$$S^2(x) = S_{\hat{y}}^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \right), \quad \text{где} \quad S_{\hat{y}}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(\breve{y}_i - a - bx_i\right)^2.$$

Значения коэффициентов $k(n, \alpha, \beta)$ и $k^*(n, \alpha, \beta)$ приведены в табл. 203 и 204 соответственно (входом в таблицы являются значения n^*). Для $n^* > 50$ справедливо приближение

$$k^*(n^*(x), \alpha, \beta) = u_{\frac{1+\beta}{2}}\left(1 + \frac{u_\alpha}{\sqrt{2n^*(x)}} + \frac{5u_\alpha^2 + 10}{12n^*(x)}\right).$$

Задача 370. *Для данных задачи 365 найти двустороннюю толерантную область, в которой с вероятностью не менее $\alpha = 0{,}95$ будет находиться не менее $\beta = 0{,}90$ всех возможных реализаций уравнения регрессии.*

Для значения $x = 1{,}2$ имеем

$$S_{\hat{y}}^2 = 30{,}447 \ (S_{\hat{y}} = 5{,}518); \quad \bar{x} = 8{,}82; \quad \sum_{i=1}^{10}(x_i - \bar{x})^2 = 313{,}216;$$

$$S_{\hat{y}}(1{,}2) = 5{,}518 \cdot \sqrt{\frac{1}{10} + \frac{(1{,}2 - 8{,}82)^2}{313{,}216}} = 2{,}948; \quad a = -4{,}9265; \quad b = 4{,}275;$$

$$\hat{y}(1{,}2) = -4{,}9276 + 4{,}275 \cdot 1{,}2 = 0{,}2024; \quad n^*(1{,}2) = \frac{10 \cdot 313{,}216}{313{,}216 + 10 \cdot (1{,}2 - 8{,}2)^2} = 3{,}5.$$

Из табл. 204 для $\alpha = 0{,}95$, $\beta = 0{,}90$ и $n^* = 3{,}5$ (используем интерполяцию) получаем $k^* = 6{,}885$.

Тогда $0{,}2035 - 2{,}948 \cdot 6{,}885 = -20{,}09 \leqslant \hat{y}(1{,}2) \leqslant 20{,}50 = 0{,}2035 + 2{,}948 \cdot 6{,}885$.

По аналогии получаем толерантные интервалы для остальных значений независимой переменной x. Результаты сведем в таблицу:

x	$\hat{y}(x)$	$S_{\hat{y}}(x)$	n^*	k^*	$\hat{y}_{\text{в}}(x)$	$\hat{y}_{\text{н}}(x)$
1,2	0,202	2,948	3,500	6,885	20,500	−20,090
2,7	6,616	2,586	4,550	4,767	18,943	−5,711
3,9	11,746	2,323	5,640	3,914	20,838	2,654
5,1	16,876	2,095	6,930	3,360	23,915	9,837
7,2	25,853	2,816	9,230	3,497	32,203	19,502
11,1	42,526	1,884	8,580	3,040	48,253	36,529
12,4	48,083	2,071	7,090	3,346	55,010	41,150
16,1	63,900	1,863	3,710	6,230	81,736	46,060
19,1	76,726	3,649	2,290	25,160	168,535	−15,083

5.3.1.3. Сравнение линейных регрессий

Сравнение двух регрессионных моделей включает в себя проверку нулевой гипотезы $H_0\colon y_1 = y_2$ о статистической неразличимости линейных регрессий $y_1 = a_1 + b_1 x$ и $y_2 = a_2 + b_2 x$. Для проверки H_0 необходимо последовательно проверить три гипотезы: о неразличимости коэффициентов a_1 и a_2 $(H_0'\colon a_1 = a_2)$; b_1 и b_2 $(H_0''\colon b_1 = b_2)$, и о равенстве остаточных дисперсий, характеризующих разброс значений y_i вокруг линии регрессии $(H_0'''\colon S_1^2 = S_2^2)$.

В первую очередь проверяется гипотеза $H_0'''\colon S_1^2 = S_2^2$. Для поверки гипотезы сначала подсчитываются соответствующие дисперсии

$$S_1^2 = \frac{1}{n_1 - 2}\sum_{i=1}^{n_1}(y_i - a_1 - b_1 x_i)^2; \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(y_i - a_2 - b_2 x_i)^2,$$

где n_1, n_2 — объемы выборок, по которым найдены регрессии y_1 и y_2 соответственно; x_{1i} $(i = 1, \ldots, n_1)$ и x_{2i} $(i = 1, \ldots, n_2)$ — значения переменной x в двух выборках.

Затем дисперсии S_1^2 и S_2^2 сравниваются между собой с помощью критерия Фишера. Если $\dfrac{S_1^2}{S_2^2} > F_\alpha$, где F_α — α-квантиль распределения Фишера с $f_1 = n_1 - 2$ и $f_2 = n_2 - 2$ степенями свободы, то гипотеза H_0''' о равенстве двух регрессий отклоняется, и регрессии признаются статистически различимыми (в числителе всегда должна быть бóльшая дисперсия).

Если гипотеза H_0''' не отклоняется, можно приступить к проверке гипотез H_0' и H_0''. Для проверки равенства угловых коэффициентов b_1 и b_2 (гипотеза H_0'') используется статистика

$$t_b = \frac{b_1 - b_2}{S^*\sqrt{\dfrac{1}{(n_1 - 1)\,S_{x_1}^2} + \dfrac{1}{(n_2 - 1)\,S_{x_2}^2}}},$$

где $$S_{\bar{x}_1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2; \quad S_{\bar{x}_2}^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2;$$

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i}; \quad \bar{x}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} x_{2i}; \quad S^* = \left\{\frac{(n_1 - 2)\,S_{x_1}^2 + (n_2 - 2)\,S_{x_2}^2}{n_1 + n_2 - 4}\right\}^{\frac{1}{2}}.$$

При справедливости нулевой гипотезы H_0'' (т. е. при $b_1 = b_2$) статистика t_b имеет распределение Стьюдента с $f = n_1 + n_2 - 4$ степенями свободы. Если

$$|t_b| \geqslant t_{\frac{1+\alpha}{2}},$$

то с доверительной вероятностью α гипотеза H_0'' отклоняется и сравниваемые угловые коэффициенты b_1 и b_2 признаются различными. Если гипотеза H_0'' не отклоняется, то проводится проверка гипотезы H_0' (проверка равенства $a_1 = a_2$). Для проверки этой гипотезы используется статистика

$$t_a = \frac{\bar{b} - \tilde{b}}{\tilde{S}},$$

$$\text{где}\quad \bar{b} = \frac{(n_1 - 1)\, S_{x_1}^2 b_1 + (n_2 - 1)\, S_{x_2}^2 b_2}{(n_1 - 1)\, S_{x_1}^2 + (n_2 - 1)\, S_{x_2}^2}; \quad \tilde{b} = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2};$$

$$\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i}; \quad \bar{y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i};$$

$$\tilde{S} = S^* \left\{ \frac{1}{(n_1 - 1)\, S_{x_1}^2 + (n_2 - 1)\, S_{x_2}^2} + \frac{1}{(\bar{x}_1 - \bar{x}_2)} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right\};$$

S^*, $S_{\bar{x}_1}^2$, $S_{\bar{x}_2}^2$ определены выше при проверке гипотезы H_0''.

При справедливости H_0' статистика t_a имеет распределение Стьюдента с $f = n_1 + n_2 - 4$ степенями свободы.

Следовательно, при справедливости $|t_a| > t_{\frac{1+\alpha}{2}}(f)$ гипотеза H_0' отклоняется.

Если все гипотезы H_0', H_0'' и H_0''' не отклоняются, то сравниваемые регрессии y_1 и y_2 признаются неразличимыми.

Задача 371. *В результате двух независимых экспериментов получены следующие результаты* ($n_1 = 10$, $n_2 = 6$):

x_{1i}:	2	4	6	9	11	16	17	20	25	31;	x_{2i}:	12	16	21	23	28	31;
y_{1i}:	9	19	22	41	49	61	69	83	98	128;	y_{2i}:	54	68	87	93	112	130.

Проверить гипотезу о статистической неразличимости регрессионных моделей, полученных по обеим выборкам при вероятности $\alpha = 0{,}95$.

Вычисляем для первой выборки

$$\sum_{i=1}^{10} x_{1i} = 141; \quad \bar{x}_1 = 14{,}1; \quad \left(\sum_{i=1}^{10} x_{1i}\right)^2 = 19881; \quad \sum_{i=1}^{10} x_{1i}^2 = 2789; \quad \sum_{i=1}^{10} y_{1i} = 589;$$

$$\bar{y}_1 = 57{,}9; \quad \sum_{i=1}^{10} x_{1i} \cdot y_{1i} = 11361;$$

$$b_1 = \frac{10 \cdot 11361 - 141 \cdot 579}{10 \cdot 2789 - 19881} = 3{,}992; \quad a = \frac{179 - 3{,}992 \cdot 14{,}1}{10} = 1{,}613;$$

$$S_{\bar{x}_1}^2 = \frac{1}{n_1 - 1} \cdot \sum_{i=1}^{n_1} (x_{1i} - \bar{x})^2 = 88{,}989 \quad (S_{\bar{x}_1} = 9{,}439).$$

Аналогично для элементов второй выборки имеем

$$\sum_{i=1}^{6} x_{2i} = 131; \quad \bar{x}_2 = 21{,}83; \quad \left(\sum_{i=1}^{6} x_{2i}\right)^2 = 17161; \quad \sum_{i-1}^{6} x_{2i}^2 = 3115;$$

$$S_{\bar{x}_2}^2 = 50{,}967 \quad (S_{\bar{x}_2} = 7{,}39);$$

$$\sum_{\imath=1}^{6} y_{2\imath} = 5446; \quad \bar{y}_2 = 90{,}667; \quad \sum_{i=1}^{6} x_{2\imath} \cdot y_{2\imath} = 12868;$$

$$b_2 = \frac{6 \cdot 12688 - 131 \cdot 544}{6 \cdot 3115 - 17161} = 3{,}887; \quad a = \frac{544 - 3{,}887 \cdot 131}{6} = 5{,}80.$$

Далее вычислим дисперсии рассеяния значений y_i вокруг линии регрессии:

$$S_1^2 = \frac{1}{n_1 - 1} \cdot \sum_{\imath=1}^{n_1} (y_{1i} - 1{,}613 - 3{,}992 \cdot x_{1\imath})^2 = \frac{1}{8} \cdot \sum_{\imath=1}^{10} (y_{1\imath} - 1{,}613 - 3{,}992 \cdot x_{1\imath})^2 = 10{,}056;$$

$$S_2^2 = \frac{1}{4} \cdot \sum_{\imath=1}^{6} (y_{2i} - 5{,}80 - 3{,}887 \cdot x_{2\imath})^2 = 7{,}0271; \quad S_1 = 3{,}171; \quad S = 2{,}651.$$

Проверим гипотезу $H_0''': S_1^2 = S_2^2$. Имеем $\dfrac{S_1^2}{S_2^2} = \dfrac{10{,}056}{7{,}0271} = 1{,}431$. Так как $\dfrac{S_1^2}{S_2^2} = 1{,}43 <$ $< F_{0,95}(8; 4) = 3{,}838$, гипотеза H_0''' принимается и дисперсии S_1^2 и S_2^2 с вероятностью $\alpha = 0{,}95$ признаются статистически неразличимыми.

Проверяем теперь гипотезу H_0'', для чего вычислим

$$S^* = \left\{ \frac{(n_1 - 2) \cdot S_1^2 + (n_2 - 2) \cdot S_2^2}{n_1 + n_2 - 4} \right\}^{\frac{1}{2}} = \left(\frac{8 \cdot 10{,}056 + 4 \cdot 7{,}0271}{12} \right)^{\frac{1}{2}} = 3{,}008.$$

Далее

$$t_b = \frac{b_1 - b_2}{S^* \cdot \sqrt{\dfrac{1}{(n_1 - 1) \cdot S_{x_1}^2} + \dfrac{1}{(n_2 - 1) \cdot S_{x_2}^2}}} = \frac{3{,}992 - 3{,}887}{3{,}008 \cdot \sqrt{\dfrac{1}{9 \cdot 88{,}988} + \dfrac{1}{5 \cdot 50{,}967}}} = 0{,}0078.$$

Из таблиц находим $t_{\frac{1+\alpha}{2}}\,(f = n_1 + n_2 - 4) = t_{0,975}\,(12) = 2{,}17$.

Так как $|t_b| = 0{,}0078 < t_{0,975}(12) = 2{,}179$, гипотеза H_0'' не отклоняется, и угловые коэффициенты b_1 и b_2 признаются неразличимыми с достоверностью $\alpha = 0{,}95$. Проверим гипотезу H_0'''. Находим:

$$\bar{b} = \frac{9 \cdot 88{,}989 \cdot 3{,}992 + 5 \cdot 50{,}967 \cdot 3{,}887}{9 \cdot 88{,}989 + 5 \cdot 50{,}967} = 3{,}967; \quad \tilde{b} = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2} = \frac{57{,}9 - 50{,}667}{14{,}1 - 21{,}83} = 4{,}239;$$

$$\tilde{S} = 3{,}008 \cdot \left\{ \frac{1}{9 \cdot 88{,}989 + 5 \cdot 50{,}967} + \frac{1}{(14{,}1 - 21{,}83)^2} \cdot \left(\frac{1}{10} + \frac{1}{6} \right) \right\}^{\frac{1}{2}} = 0{,}2212;$$

$$t_a = \frac{3{,}967 - 4{,}239}{0{,}2212} = -1{,}23.$$

Сравнивая $|t_a|$ с $t_{0,975}(12)$, имеем $|t_a| = 1{,}230 < t_{0,975}(12) = 2{,}170$.

Следовательно, и гипотеза H_0''' не отклоняется. Таким образом, сравнением эмпирических уравнений регрессий убеждаемся, что они с достоверностью не менее 0,95 статистически эквивалентны.

5.3.1.4. Некоторые специальные задачи линейного регрессионного анализа

В этом разделе мы рассмотрим некоторые частные задачи, возникающие в линейном регрессионном анализе. В связи с их специфичностью они не рассматриваются в классических курсах регрессионного анализа.

5.3.1.4.1. Оценка вершины кусочно-ломаной линии регрессии

Предположим, что имеются две пересекающиеся линии регрессии и требуется определить доверительную зону для точки их пересечения.

Алгоритм получения такой оценки включает в себя ряд последовательных операций. Сначала строится двусторонний α'-доверительный интервал для двух

угловых коэффициентов b_1 и b_2 регрессии рассматриваемых прямых $y_1 = a_1 + b_1 x$ и $y_2 = a_2 + b_2 x$ соответственно. С помощью методов, изложенных в разделе 5.3.1.2.1.1, находим [106]

$$\hat{b}_1 - t_{\frac{1+\alpha'}{2}} S_{\beta 1} \leqslant b_1 \leqslant \hat{b}_1 + t_{\frac{1+\alpha'}{2}} S_{\beta 1}; \quad \hat{b}_2 - t_{\frac{1+\alpha'}{2}} S_{\beta 2} \leqslant b_2 \leqslant \hat{b}_2 + t_{\frac{1+\alpha'}{2}} S_{\beta 2},$$

где t_γ — γ-квантиль распределения Стьюдента с $f = n - 2$ степенями свободы (n — объем выборки, будем считать его одинаковым для обеих выборок); $\hat{b}_1$, $\hat{b}_2$ — оценки наименьших квадратов соответственно для угловых коэффициентов b_1 и b_2, равные (см. раздел 5.3.1.1.1)

$$\hat{b}_1 = \frac{n \sum_{i=1}^{n} x_{1i} \bar{y}_{1i} - \sum_{i=1}^{n} x_{1i} \sum_{i=1}^{n} \bar{y}_{1i}}{n \sum_{i=1}^{n} x_{1i}^2 - \left(\sum_{i=1}^{n} x_{1i} \right)^2}, \quad \hat{b}_2 = \frac{n \sum_{i=1}^{n} x_{2i} \bar{y}_{2i} - \sum_{i=1}^{n} x_{2i} \sum_{i=1}^{n} \bar{y}_{2i}}{n \sum_{i=1}^{n} x_{2i}^2 - \left(\sum_{i=1}^{n} x_{2i} \right)^2},$$

где $\bar{y}_{1i} = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{1j}$; $\bar{y}_{2i} = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{2j}$;

$$S_{\beta 1} = \frac{S_1}{S_{\bar{x}_1} \sqrt{n-1}},$$

где $S_{\bar{x}_1}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2$; $\bar{x}_1 = \frac{1}{n} \sum_{i=1}^{n} x_{1i}$; $S_1^2 = \frac{1}{n-2} \left(\bar{y}_{1i} - \hat{a}_1 - \hat{b}_1 x_{1i} \right)^2$;

$$S_{\beta 2} = \frac{S_2}{S_{\bar{x}_2} \sqrt{n-1}},$$

где $S_{\bar{x}_2}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2$; $\bar{x}_2 = \frac{1}{n} \sum_{i=1}^{n} x_{2i}$; $S_2^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left(\bar{y}_{2i} - \hat{a}_2 - \hat{b}_2 x_{2i} \right)^2$.

Оценки $\hat{a}_1$ и $\hat{a}_2$ равны

$$\hat{a}_1 = \frac{\sum_{i=1}^{n} \bar{y}_{1i} - \hat{b}_1 \sum_{i=1}^{n} x_{1i}}{n}; \quad \hat{a}_2 = \frac{\sum_{i=1}^{n} \bar{y}_{2i} - \hat{b}_2 \sum_{i=1}^{n} x_{2i}}{n}.$$

Тогда двусторонние α-доверительные интервалы для средних $\bar{y}_1 = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_{1i}$ и $\bar{y}_2 = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_{2i}$ будут иметь вид:

$$\bar{y}_1 - t_{\frac{1+\alpha''}{2}} \frac{S_{\bar{y}_1}}{\sqrt{n}} \leqslant M(\bar{y}_1) \leqslant \bar{y}_1 + t_{\frac{1+\alpha''}{2}} \frac{S_{\bar{y}_1}}{\sqrt{n}}; \quad \bar{y}_2 - t_{\frac{1+\alpha''}{2}} \frac{S_{\bar{y}_2}}{\sqrt{n}} \leqslant M(\bar{y}_2) \leqslant \bar{y}_2 + t_{\frac{1+\alpha''}{2}} \frac{S_{\bar{y}_2}}{\sqrt{n}},$$

где

$$S_{\bar{y}_1}^2 = \frac{(m_i - 1) S_{1i}^2 + \ldots + (m_n - 1) S_{1n}^2}{\sum_{i=1}^{n} m_i - n}; \quad S_{1i}^2 = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (y_{1ij} - \bar{y}_{1i})^2;$$

$$S_{\bar{y}_2}^2 = \frac{(m_i - 1) S_{2i}^2 + \ldots + (m_n - 1) S_{2n}^2}{\sum_{i=1}^{n} m_i - n}; \quad S_{2i}^2 = \frac{1}{m_i - 1} \sum_{j=1}^{n} (y_{2ij} - \bar{y}_{2i});$$

m_i — число параллельных наблюдений над y_1 и y_2 на уровне x_i (будем считать для простоты, что $m_i = \text{const}$ для всех x_i, y_1 и y_2).

Легко видеть, что S_{1i}^2 и S_{2i}^2 — это дисперсии, характеризующие рассеяние величин y_1 и y_2 на каждом уровне x_i; $S_{\bar{y}_1}^2$ и $S_{\bar{y}_2}^2$ — средневзвешенные дисперсии величин $\bar{y}_1$ и $\bar{y}_2$ по всем x_i.

Следует помнить, что для оценки дисперсий $S_{\bar{y}_1}^2$ и $S_{\bar{y}_2}^2$ необходимы параллельные наблюдения над y на каждом уровне x. В нашем случае $t_{\frac{1+\alpha''}{2}}$ — соответствующая квантиль распределения Стьюдента с $f = \sum_{i=1}^{n} m_i - 1$ степенями свободы.

Теперь через каждую точку с координатами $(\bar{x}_1, \bar{y}_1^{\text{н}})$ и $(\bar{x}_1, \bar{y}_1^{\text{в}})$ проведем по две прямые с угловыми коэффициентами $b_1^{\text{н}}$ и $b_1^{\text{в}}$ ($\bar{y}_1^{\text{н}}, \bar{y}_1^{\text{в}}, b_1^{\text{н}}$ и $b_1^{\text{в}}$ — соответственно нижние и верхние границы доверительных интервалов). То же самое выполняем и для второй регрессии — т. е. проведем прямые через точки $(\bar{x}_2, \bar{y}_2^{\text{н}})$ и $(\bar{x}_2, \bar{y}_2^{\text{в}})$.

Через максимальную область, охватываемую прямыми

$$\begin{aligned} y &= \bar{y}_1^{\text{н}} + b_1^{\text{н}}(x - \bar{x}_1); \quad y = \bar{y}_1^{\text{н}} + b_1^{\text{в}}(x - \bar{x}_1) \quad \text{и} \\ y &= \bar{y}_1^{\text{в}} + b_1^{\text{н}}(x - \bar{x}_1); \quad y = \bar{y}_1^{\text{в}} + b_1^{\text{в}}(x - \bar{x}_1), \end{aligned}$$

будет с вероятностью $\beta = \alpha'\alpha''$ проходить линия регрессии $y_1 = a_1 + b_1 \cdot x$.

По аналогии через максимальную область, охватываемую прямыми

$$\begin{aligned} y &= \bar{y}_2^{\text{н}} + b_2^{\text{н}}(x - \bar{x}_2); \quad y = \bar{y}_2^{\text{н}} + b_2^{\text{в}}(x - \bar{x}_2) \quad \text{и} \\ y &= \bar{y}_2^{\text{в}} + b_2^{\text{н}}(x - \bar{x}_2); \quad y = \bar{y}_2^{\text{в}} + b_2^{\text{в}}(x - \bar{x}_2), \end{aligned}$$

будет с вероятностью $\beta = \alpha'\alpha''$ проходить линия регрессии $y_2 = a_2 + b_2 x$.

Пересечение этих двух областей представляет собой неправильный многоугольник, который и служит доверительной областью для точки пересечения прямых регрессии при доверительной вероятности $\gamma = \beta^2$.

Задача 372. *В результате наблюдений были получены две выборки результатов, представленных в таблице. Необходимо найти 90%-ю доверительную область для точки пересечения линий регрессии, порождаемых двумя выборками.*

x_{1i}	y_{1i}	$\bar{y}_{1i}$	S_{1i}^2	x_{2i}	y_{2i}	$\bar{y}_{2i}$	S_{2i}^2
2	19, 21, 16	18,67	6,33	0,8	15; 13; 17,1	15,03	4,203
3	26, 21, 38	28,33	76,33	1,2	17; 16,1; 18, 2	17,10	1,110
5	41, 38, 49	42,67	32,33	2,0	18; 17,1; 19, 4	18,17	1,343
9	70, 61, 82	71,00	111,00	3,0	21; 19,8; 23, 1	21,30	2,750
11	84, 71, 99	84,67	196,33	4,8	27,5; 26; 29	27,50	2,250
14	105, 91, 116	104,00	157,00	6,9	34; 31; 37	34,00	9,000
17	125, 109, 141	125,00	256,00	8,2	38; 36; 41	38,33	6,333
20	145, 131, 160	145,33	210,33	10,1	44; 41,2; 46, 1	43,76	6,043
25	180, 161, 201	180,67	400,33	12,0	50; 39; 62	50,33	132,330
31	210, 191, 231	210,67	400,33	14,0	56; 41; 69	55,33	196,330

Имеем $\gamma = \beta^2 = 0{,}90$ и $\beta = \sqrt{0{,}90} \approx 0{,}95$. Выбираем $\alpha' = \alpha'' = \alpha = \sqrt{\beta} = \sqrt{0{,}95} \approx 0{,}975$. Выполняя необходимые расчеты, получаем

$$\sum_{i=1}^{10} x_{1i} = 137; \quad \sum_{i=1}^{10} x_{1i}^2 = 2711; \quad \left(\sum_{i=1}^{10} x_{1i}\right)^2 = 18769; \quad \sum_{i=1}^{10} \bar{y}_{1i} = 1011{,}01;$$

$$\sum_{i=1}^{10} x_{1i}\cdot\bar y_{1i}=19441{,}17;\quad \sum_{i=1}^{10} x_{2i}=63;\quad \sum_{i=1}^{10} x_{2i}^2=594{,}98;\quad \left(\sum_{i=1}^{10} x_{2i}\right)^2=3969;$$

$$\sum_{i=1}^{10}\bar y_{2i}=320{,}185;\quad \sum_{i=1}^{10} x_{2i}\cdot\bar y_{2i}=2634{,}246.$$

Находим оценки коэффициентов регрессии:

$$\hat b_1=\frac{10\cdot 19441{,}17-137\cdot 1011{,}01}{10\cdot 2711-137^2}=6{,}702;\quad \hat a_1=\frac{1011{,}01-6{,}702\cdot 137}{10}=9{,}2836;$$

$$\hat b_2=\frac{10\cdot 2634{,}246-63\cdot 320{,}85}{10\cdot 594{,}98-3969}=3{,}094;\quad \hat a_2=\frac{320{,}85-3{,}094\cdot 63}{10}=12{,}593.$$

Вычисляем дисперсии

$$S_1^2=\frac{1}{n-2}\cdot\sum_{i=1}^{n}(\bar y_{1i}-a_1-b_1\cdot x_{1i})^2=10{,}6703\ (S_1=3{,}2665);$$

$$S_2^2=\frac{1}{n-2}\cdot\sum_{i=1}^{n}(\bar y_{2i}-a_2-b_2\cdot x_{2i})^2=0{,}2737\ (S_2=0{,}52317);$$

$$S_{\bar x_1}^2=\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_{1i}-\bar x_1)^2=92{,}678\ (S_{x_1}=9{,}627);$$

$$S_{\bar x_2}^2=\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_{2i}-\bar x_2)^2=22{,}009;\ (S_{x_2}=4{,}691);$$

$$S_{\beta 1}=\frac{S_1}{S_{\bar x_1}\cdot\sqrt{n-1}}=\frac{3{,}2665}{9{,}627\cdot 3}=0{,}1131;\quad S_{\beta 2}=\frac{S_2}{S_{\bar x_2}\cdot\sqrt{n-1}}=\frac{0{,}52317}{4{,}691\cdot 3}=0{,}03717.$$

Теперь из табл. 118 (или аппроксимацией из раздела 1.1.9 для $\alpha'=0{,}975$) находим (интерполируя) $t_{(1+0{,}975)/2}=t_{0{,}9875}(f=8)=2{,}798$.

Для двустороннего доверительного интервала имеем

$$b_1^{\text{н}}=b_1-t_{0{,}9875}\cdot S_{\beta 1}=6{,}702-2{,}798\cdot 0{,}1131=6{,}385;$$
$$b_1^{\text{в}}=b_1+t_{0{,}9875}\cdot S_{\beta 1}=6{,}702+2{,}798\cdot 0{,}1131=7{,}018;$$
$$b_2^{\text{н}}=b_2-t_{0{,}9875}\cdot S_{\beta 2}=3{,}094-2{,}798\cdot 0{,}03717=2{,}990;$$
$$b_2^{\text{в}}=b_2+t_{0{,}9875}\cdot S_{\beta 2}=3{,}094+2{,}798\cdot 0{,}03717=3{,}198.$$

Теперь вычислим дисперсии ($m_i=3$):

$$S_{\bar y_1}^2=\frac{m_i-1}{\sum_{i=1}^{n} m_i-n}\cdot\sum_{i=1}^{n}S_{1i}^2=\frac{2\cdot 1846{,}313}{20}=184{,}6313\ \ (S_{\bar y_1}=13{,}588)\,;$$

$$S_{\bar y_2}^2=\frac{m_i-1}{\sum_{i=1}^{n} m_i-n}\cdot\sum_{i=1}^{n}S_{2i}^2=\frac{2\cdot 361{,}692}{20}=36{,}1692\ \ (S_{\bar y_2}=6{,}014)\,.$$

Так как оценки $S_{\bar y_1}^2$ и $S_{\bar y_2}^2$ используют все $\sum_{i=1}^{n} m_i-n=20$ степеней свободы, то

$$t_{\frac{1+\alpha''}{2}}=t_{\frac{1+0{,}975}{2}}=t_{0{,}9875}(f=20-1=19)=2{,}465.$$

Далее вычисляем

$$\bar y_1^{\text{н}}=\bar y_1-t_{0{,}9875}\cdot\frac{S_{y_1}}{\sqrt n}=\frac{1}{10}\cdot\sum_{i=1}^{10}y_{1i}-2{,}465\cdot\frac{13{,}588}{\sqrt{10}}=101{,}101-\frac{2{,}465\cdot 13{,}588}{\sqrt{10}}=90{,}509;$$

$$\bar y_1^{\text{в}}=\bar y_1+t_{0{,}9875}\cdot\frac{S_{y_1}}{\sqrt n}=101{,}101+2{,}465\cdot\frac{13{,}588}{\sqrt{10}}=111{,}693;$$

$$\bar{y}_2^{\text{н}} = \bar{y}_2 - t_{0,9875} \cdot \frac{S_{y2}}{\sqrt{n}} = \frac{1}{10} \cdot \sum_{i=1}^{10} y_{2_i} - 2{,}645 \cdot \frac{6{,}014}{\sqrt{10}} = 32{,}085 - \frac{2{,}645 \cdot 6{,}014}{\sqrt{10}} = 27{,}397;$$

$$\bar{y}_2^{\text{в}} = \bar{y}_2 + t_{0,9875} \cdot \frac{S_{y2}}{\sqrt{n}} = 32{,}085 + 2{,}465 \cdot \frac{6{,}014}{\sqrt{10}} = 36{,}773.$$

Максимальная область, охватываемая прямыми

$$y = 90{,}509 + 6{,}385 \cdot (x - 13{,}7)\,; \quad y = 90{,}509 + 7{,}018 \cdot (x - 13{,}7)\,;$$
$$y = 111{,}693 + 6{,}385 \cdot (x - 13{,}7)\,; \quad y = 111{,}693 + 7{,}018 \cdot (x - 13{,}7)\,,$$

содержит линию регрессии $\hat{y}_1 = 9{,}2836 + 6{,}702 \cdot x$ с вероятностью

$$\beta = \alpha' \cdot \alpha'' = 0{,}9875 \cdot 0{,}9875 = 0{,}975.$$

По аналогии, максимальная область, охватываемая прямыми

$$y = 27{,}397 + 2{,}990 \cdot (x - 6{,}3)\,; \quad y = 27{,}397 + 3{,}198 \cdot (x - 6{,}3)\,;$$
$$y = 36{,}773 + 2{,}990 \cdot (x - 6{,}3)\,; \quad y = 36{,}773 + 3{,}198 \cdot (x - 6{,}3)\,,$$

содержит линию регрессии $\hat{y}_2 = 12{,}593 + 3{,}094 \cdot x$ с вероятностью

$$\beta = \alpha' \cdot \alpha'' = 0{,}9875 \cdot 0{,}9875 = 0{,}975.$$

Поясним максимальную область для первой прямой регрессии — она ограничивается:

— для $x > 13{,}7$ прямыми $y_1 = 111{,}93 + 7{,}018 \cdot (x - 13{,}7)$ и $y_1' = 90{,}509 + 6{,}385 \cdot (x - 13{,}7)$;

— для $x < 13{,}7$ прямыми $\tilde{y}_1 = 111{,}93 + 6{,}385 \cdot (x - 13{,}7)$ и $\tilde{y}_1' = 90{,}509 + 7{,}018 \cdot (x - 13{,}7)$.

Аналогично, для второй линии регрессии имеем максимальную область:

— для $x > 6{,}3$ между линиями $y_2 = 36{,}773 + 3{,}198 \cdot (x - 6{,}3)$; $y_2' = 27{,}397 + 2{,}99 \times (x - 6{,}3)$;

— для $x < 6{,}3$ между линиями $\tilde{y}_2 = 36{,}773 + 2{,}99 \cdot (x - 6{,}3)$; $\tilde{y}_2' = 27{,}397 + 3{,}198 \times (x - 6{,}3)$.

Из анализа максимальных областей линий регрессии видим, что они пересекаются при $x < 6{,}3$.

Теперь осталось найти четыре точки пересечения четырех линий, ограничивающих максимальные области пересекающихся линий регрессии.

Для $x < 6{,}3$ имеем следующие линии:

$$\tilde{y}_1 = 111{,}693 + 6{,}385 \cdot (x - 13{,}7)\,; \quad \tilde{y}_1' = 90{,}509 + 7{,}018 \cdot (x - 13{,}7)\,;$$
$$\tilde{y}_2 = 36{,}773 + 2{,}99 \cdot (x - 6{,}3)\,; \quad \tilde{y}_2' = 27{,}397 + 3{,}198 \cdot (x - 6{,}3)\,,$$

или

$$\tilde{y}_1 = 24{,}2185 + 6{,}385 \cdot x; \quad \tilde{y}_1' = -5{,}6376 + 7{,}018 \cdot x;$$
$$\tilde{y}_2 = 17{,}936 + 2{,}99 \cdot x; \quad \tilde{y}_2' = 7{,}2496 + 3{,}198 \cdot x.$$

Находим попарные точки пересечения прямых $(\tilde{y}_1, \tilde{y}_2')$, $(\tilde{y}_1, \tilde{y}_2)$, $(\tilde{y}_1', \tilde{y}_2)$, $(\tilde{y}_1, \tilde{y}_2')$, попадающие в область определения $x < 6{,}3$. Прямые $\tilde{y}_1$ и $\tilde{y}_2'$ пересекаются в точке $x = -5{,}32$, прямые $\tilde{y}_1$ и $\tilde{y}_2$ в точке $x = -1{,}85$, прямые $\tilde{y}_1'$ и $\tilde{y}_2$ в точке $x = 5{,}85$, прямые $\tilde{y}_1'$ и $\tilde{y}_2'$ в точке $x = 3{,}37$.

Таким образом, мы имеем для описания области пересечения прямых регрессии четыре точки с координатами ($x = -5{,}32$; $y = -9{,}75$), ($x = -1{,}85$; $y = 12{,}4$), ($x = 5{,}85$; $y = 35{,}43$), ($x = 3{,}37$; $y = 18{,}01$).

С вероятностью $\gamma = \beta^2 = (0{,}95 \cdot 0{,}95)^2 = 0{,}90$ в области, ограниченной этими точками, находится точка пересечения двух линий регрессии. Отметим, что искомая область может быть описана и бо́льшим количеством точек пересечения.

5.3.1.4.2. Определение объема испытаний для получения заданной точности оценки коэффициента регрессии

Задача рассмотрена в [598] и формулируется следующим образом. Для известной заранее стандартной ошибки отдельного наблюдения $y_i(\sigma)$ необходимо так спланировать наблюдения, чтобы ошибка в определении углового коэффициента

регрессии методом наименьших квадратов при заданной доверительной вероятности α не превышала Δb (при этом желательно, чтобы число наблюдений n и интервал изменения независимой переменной $x(l)$ были минимальными).

При наиболее неблагоприятном экстремальном расположении наблюдений (половина на одном конце интервала, половина на другом) минимально допустимое соотношение между числом наблюдений n и длиной интервала l для достижения заданной точности Δb в определении коэффициента регрессии имеет вид [598]

$$l \geqslant \frac{\sigma}{\Delta b} k(\alpha, n), \quad \text{где} \quad k(\alpha, n) = \begin{cases} 2t_\alpha(n)\sqrt{\dfrac{n-1}{n(n-2)}}, & n = 2m; \\ 2t_\alpha(n)\sqrt{\dfrac{n}{(n+1)(n-2)}}, & n = 2m-1 \end{cases}$$

— коэффициент, зависящий от доверительной вероятности α и числа наблюдений n (его значения приведены в табл. 236).

Таблица 236

Значения коэффициента $k(\alpha, n)$ [598]

n	Доверительная вероятность α			n	Доверительная вероятность α		
	0,90	0,95	0,99		0,90	0,95	0,99
3	10,93	22,01	55,11	16	0,91	1,11	1,36
4	3,58	18,49	48,58	17	0,88	1,07	1,30
5	2,48	3,35	4,78	18	0,85	1,03	1,25
6	1,94	2,54	3,42	19	0,82	1,00	1,21
7	1,69	2,15	2,82	20	0,79	0,97	1,17
8	1,48	1,87	2,40	21	0,77	0,94	1,14
9	1,36	1,70	2,15	22	0,75	0,91	1,10
10	1,25	1,55	1,94	23	0,73	0,89	1,08
11	1,17	1,44	1,80	24	0,72	0,86	1,05
12	1,10	1,35	1,68	25	0,70	0,85	1,03
13	1,04	1,28	1,58	26	0,68	0,83	1,00
14	0,99	1,21	1,49	30	0,63	0,76	0,92
15	0,95	1,16	1,42	60	0,48	0,52	0,62

По заданным σ и Δb можно найти соотношения между l и Δb, между l и n и, задаваясь одним, выбирать другое.

Задача 373. *Известно, что стандартная ошибка измерений y равна $\sigma = 5{,}27$. Необходимо определить:*

— количество наблюдений, обеспечивающих ошибку оценки коэффициента регрессии не более $\Delta b = 3{,}2$ на интервале изменения x длиной $l = 10$ при доверительной вероятности $\alpha = 0{,}95$;

— наименьшую длину интервала изменения переменной x, обеспечивающую точность оценки $\Delta b = 1{,}1$ с помощью $n = 25$ наблюдений при доверительной вероятности $\alpha = 0{,}90$.

Отвечаем на первый вопрос. Имеем $l = \dfrac{\sigma}{\Delta b} \cdot k(\alpha, n)$ и для $l = 10$, $\alpha = 0{,}95$, $\sigma = 5{,}27$ и $\Delta b = 3{,}2$ получаем $k(0{,}95; n) = 10 \cdot 3{,}2/5{,}27 = 6{,}07$. Из табл. 236 имеем $n = 5$.

Отвечаем на второй вопрос. Для $n = 25$, $\Delta b = 1{,}1$, $\alpha = 0{,}90$ из табл. 236 имеем $k(0{,}90; 25) = 0{,}70$. Тогда $l = \dfrac{5{,}27 \cdot 0{,}7}{1{,}1} = 3{,}35$.

5.3.2. Множественная линейная регрессия

Ранее рассматривались методы оценки и анализа линейной регрессии $y = f(x)$. Если случайная величина y может зависеть одновременно от двух и более переменных, возникает задача оценки и анализа множественной регрессии $y = f(x_1, x_2, \ldots, x_k)$, где x_j $(j = 1, 2, \ldots, k)$ — независимые переменные.

Рассмотрим в качестве примера регрессию $y = \alpha + \beta_1 x_1 + \beta_2 x_2$ для случая двух независимых переменных x_1 и x_2. Для большего числа переменных оценки могут быть получены по аналогии. Для нашего случая оценки коэффициентов β_1 , β_2 и α имеют вид:

$$b_1 = \frac{S(x_2^2)\,S(x_1, y) - S(x_1x_2)\,S(x_2y)}{S(x_1^2)\,S(x_2^2) - [S(x_1x_2)]^2}; \quad b_2 = \frac{S(x_1^2)\,S(x_2y) - S(x_1x_2)\,S(x_1y)}{S(x_1^2)\,S(x_2^2) - [S(x_1x_2)]^2};$$

$$a = \bar{y} - b_1x_1 - b_2x_2, \quad \text{где}$$

$$S(x_j^2) = \sum_{i=1}^{n}(x_{ji} - \bar{x}_j)^2, \quad j = 1, 2; \qquad S(x_jy) = \sum_{i=1}^{n}(x_{ji} - \bar{x}_j)(y_i - \bar{y}), \quad j = 1, 2;$$

$$S(x_1x_2) = \sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2); \quad S(y^2) = \sum_{i=1}^{n}(y_i - \bar{y})^2.$$

Дисперсия, характеризующая разброс значений y_i вокруг линии регрессии, равна

$$S^2 = \frac{1}{n-3}\sum_{i=1}^{n}(y_i - a - b_1x_1 - b_2x_2)^2 = \frac{1}{n-3}\left[S(y^2) - b_1S(x_1y) - b_2S(x_2y)\right].$$

Проверка гипотез $H_0'\colon b_1 = b_{10}$ и $H_0''\colon b_2 = b_{20}$ (b_{10}, b_{20} — некоторые заранее заданные значения) производится с помощью статистик

$$t_1 = \frac{b_1 - b_{10}}{S_{b_1}}; \quad t_2 = \frac{b_2 - b_{20}}{S_{b_2}}, \quad \text{где}$$

$$S_{b_1} = S\left\{\frac{S(x_2^2)}{S(x_1^2)\,S(x_2^2) - [S(x_1x_2)]^2}\right\}^{\frac{1}{2}}; \quad S_{b_2} = S\left\{\frac{S(x_1^2)}{S(x_1^2)\,S(x_2^2) - [S(x_1x_2)]^2}\right\}^{\frac{1}{2}}.$$

Статистики t_1 и t_2 при справедливости гипотез H_0' и H_0'' имеют распределение Стьюдента с $f = n - 3$ степенями свободы. Отсюда: если $|b_1| > t_{\frac{1+\alpha}{2}} S_{b_1}$ и $|b_2| > t_{\frac{1+\alpha}{2}} S_{b_2}$, то коэффициенты b_1 и b_2 признаются значимыми.

Сама регрессия признается значимой, если $\dfrac{S^2}{S_{y_i}^2} < F_\alpha$, где $S_{y_i}^2$ — дисперсия значений y_i, оцененная ранее или оцениваемая по дублируемым наблюдениям (для F_α используются степени свободы $f_1 = n - 3$ и $f_2 = m - 1$, где m — объем выборки, по которой производилась оценка $S_{y_i}^2$).

Для обнаружения и устранения выбросов при множественной линейной регрессии применяются статистики Прескотта–Лунда (см. раздел 5.3.1.2.2.3.3) с использованием соответствующего значения k (в нашем случае $k = 3$).

Для большего числа независимых переменных расчеты и анализ множественной регрессии существенно усложняются. Однако в настоящее время разработаны специальные методы планирования регрессионных экспериментов, позволяющие упростить оценки коэффициентов регрессии и сократить число необходимых экспериментов.

Некоторые из них будут рассмотрены далее, однако в целом этот обширнейший раздел современной прикладной статистики останется за бортом нашего труда, так как он заслуживает самостоятельного скрупулезного и пространного изложения. В самых общих чертах (на уровне элементарного введения) мы рассмотрим эти вопросы в разделе, посвященном методам активного планирования эксперимента.

5.3.3. Нелинейный регрессионный анализ

Если линейное уравнение регрессии неадекватно описывает имеющийся статистический материал, необходимо найти нелинейное уравнение регрессии, удовлетворительно описывающее истинную зависимость. Различные приемы нахождения таких зависимостей рассматриваются ниже.

5.3.3.1. Линеаризация нелинейной модели заменой переменных

Смысл метода ясен из названия — нелинейное уравнение преобразуется в линейное переходом от исследуемых переменных к новым переменным. Некоторые из таких преобразований приведены в табл. 230. Обработка результатов наблюдений, вычисление регрессии и ее статистический анализ для линейно преобразованного уравнения выполняются методами линейного регрессионного анализа, рассмотренными выше. Следует, однако, помнить исходные предпосылки регрессионного анализа, основная из которых — нормальность распределения изучаемых случайных величин — должна теперь быть справедлива не для самой случайной величины, а для ее функционального преобразования, что, к сожалению, далеко не всегда выполняется. Подбор необходимого линеаризующего преобразования выполняется первоначально „на глаз" графически. Поясним методику вычисления нелинейной регрессии примером.

Задача 374. *В результате исследования зависимости между случайными величинами x и y получены следующие результаты ($n = 10$):*

x_i:	1	2	3	4	5	6	7	8	9	10
y_i:	81	19	9	8,5	7	6,6	5	5,8	5,6	5,4.

Необходимо найти регрессию $y = f(x)$ и провести ее статистический анализ.

Предварительным анализом убеждаемся, что можно использовать линеаризующее преобразование для функции типа $y = ae^{\frac{b}{x}}$.

Обозначим $y^* = \ln y$ и $x^* = \dfrac{1}{x}$. Тогда будем искать линейную регрессию $y^* = a^* + b^* x^*$. Для x^* и y^* имеем ряд

$x^* \to 1$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{7}$	$\frac{1}{8}$	$\frac{1}{9}$	$\frac{1}{10}$
$y^* \to 4{,}39$	2,94	2,20	2,14	1,94	1,89	1,79	1,76	1,72	1,69

Находим

$$\sum_{i=1}^{10} x_i^* = 2{,}92896; \quad \left(\sum_{i=1}^{10} x_i^*\right)^2 = 8{,}578806;$$

$$\sum_{i=1}^{10} {x_i^*}^2 = 1{,}549768; \quad \sum_{i=1}^{10} y_i^* = 22{,}46; \quad \sum_{i=1}^{10} x_i^* \cdot y_i^* = 8{,}6671587.$$

Тогда

$$b^* = \frac{10 \cdot 8{,}6671587 - 2{,}92896 \cdot 22{,}46}{10 \cdot 1{,}549768 - 8{,}578806} = 3{,}019; \quad a^* = \frac{22{,}46 - 3{,}019 \cdot 2{,}92896}{10} = 1{,}361.$$

Следовательно, искомая регрессия имеет вид

$$y^* = 1{,}361 + 3{,}019 \cdot x^*, \quad \text{или} \quad \ln y = 1{,}361 + 3{,}019 \cdot \frac{1}{x}; \quad y = 3{,}9 \cdot e^{\frac{3{,}019}{x}}.$$

Проверяем значимость коэффициентов регрессии. Имеем

$$S^2 = \frac{1}{n-2} \cdot \sum_{i=1}^{n} (y_i^* \cdot 1{,}361 - 3{,}019 \cdot x^*)^2 = 0{,}004 \ \ (S = 0{,}067)\,;$$

$$S_{x^*}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i^* - \bar{x}^*)^2 = 0{,}0768 \ \ (S_{x^*} = 0{,}2772)\,; \quad S_{b^*} = \frac{S}{S_{x^*} \cdot \sqrt{n-1}} = 0{,}0805.$$

Из табл. 118 для $\alpha = 0{,}95$ имеем $t_{(1+0{,}95)/2}(f = 8) = t_{0{,}975}(8) = 2{,}31$.

Так как $|b^*| = 3{,}019 > t_{0{,}975}(8) \cdot S_{b^*} = 2{,}31 \cdot 0{,}0805 = 0{,}186$, коэффициент регрессии признается значимым. Для дисперсии коэффициента a^* имеем

$$S_{a^*} = S \cdot \left\{ \frac{1}{n} + \frac{\bar{x^*}^2}{(n-1) \cdot S_{x^*}^2} \right\}^{\frac{1}{2}} = 0{,}067 \cdot \left\{ \frac{1}{10} + \frac{0{,}292896^2}{9 \cdot 0{,}0768} \right\}^{\frac{1}{2}} = 0{,}0317.$$

Тогда $|a^*| = 1{,}361 > t_{0{,}975}(8) \cdot S_{a^*} = 2{,}31 \cdot 0{,}0317 = 0{,}073$, что также приводит к выводу о значимости коэффициента a^* с достоверностью $\alpha = 0{,}95$.

5.3.3.2. Полиномиальная нелинейная регрессия (полиномы Чебышева)

При невозможности замены переменных приходится искать оценки нелинейной регрессии непосредственно по экспериментальным данным.

Как правило, любая нелинейная регрессия может быть представлена полиномом вида $y = \sum_{i=0}^{n} a_i x^i$. Если степень полинома k выбрана заранее, то оценки коэффициентов $a_0, a_1, \ldots, a_k$ находятся непосредственно методом наименьших квадратов. При неизвестной степени аппроксимирующего полинома уравнение регрессии ищется путем последовательных уточнений (последовательного повышения степени полинома). Критерием для прекращения процедуры уточнения степени полинома является величина остаточной дисперсии

$$S_k^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{k} a_j x_i^j \right)^2.$$

Если $S_{k+1}^2 > S_k^2$, то в качестве регрессии принимается полином степени k. Значимость дисперсий S_{k+1}^2 и S_k^2 проверяется F-критерием Фишера при $f_1 = n - k$ и $f_2 = n - k - 1$ степенях свободы.

При увеличении степени полинома необходимо пересчитывать все коэффициенты регрессии. Расчеты упрощаются, если значения x_i являются равноотстоящими ($x_{i+1} - x_i = \text{const}$). Тогда уравнение регрессии k-го порядка может быть представлено в виде $y = b_0 P_0(x) + b_1 P_1(x) + \ldots + b_k P_k(x)$, где $P_k(x)$ — полиномы Чебышева, определяемые соотношениями

$$P_0(x) = 1; \quad P_1(x) = x - \frac{n+1}{2}; \quad P_k(x) = P_1(x)\, P_{k-1}(x) - \frac{k^2(n^2 - k^2)}{4(4k^2 - 1)} P_{k-2}(x)\,.$$

Коэффициенты регрессии вычисляются по формуле

$$b_k = \frac{\sum\limits_{i=1}^{n} y_i P_k(x_i)}{\sum\limits_{i=1}^{n} P_k^2(x_i)}.$$

Если $x^* = \dfrac{x - x_1}{x_2 - x_1} + 1$ $(x_2 - x_1 = ... = x_{i+1} - x_i = \ldots = x_n - x_{n-1} = \text{const})$, то $x_i^* = i$ и уравнение регрессии принимает вид

$$y = b_0 P_0(i) + b_1 P_1(i)(x - \bar{x}) + \ldots + b_k P_k(i)(x - \bar{x})^k + \ldots, \quad i = 1, 2, \ldots, n,$$

где

$$b_k = \frac{\sum_{i=1}^{n} y_i P_k(i)}{\sum_{i=1}^{n} P_k^2(i)}.$$

Значения

$$\sum_{i=1}^{n} P_k^2(i) = \frac{(k!) \cdot n \cdot (n^2 - 1) \cdot (n^2 - 4) \cdot s\,(n^2 - k^2)}{[(2k-1)!!]^2 \cdot 2^{2^k} \cdot (2k+1)},$$

$$\text{где } (2k-1)!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2k-1),$$

можно заранее табулировать для разных n и k, а значения $P_j(i)$ также вычислить, пользуясь приведенными формулами. Для некоторых значений k и n полиномы Чебышева $P_j(i)$ и $\sum_{i=1}^{n} P_k^2(i)$ приведены в табл. 237. В ней учтено, что $P_k(i) = -P_k(n+1-i)$ для нечетных k, и $P_k(i) = P_k(n+1-i)$ для четных k. Остаточная дисперсия уравнения регрессии равна

$$\tilde{S}_k^2 = \frac{1}{n-k-1} S_k^2, \quad \text{где} \quad S_k^2 = S_{k-1}^2 - b_k^2 \sum_{i=1}^{n} P_k^2(i)\,; \quad S_0^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2.$$

Проверка гипотез о значениях коэффициентов регрессии осуществляется с помощью статистик

$$t_j = \frac{(b_j - b_{j0})\left[\sum_{i=1}^{n} P_k^2(i)\right]^2}{S_k},$$

имеющих распределение Стьюдента с $f = n - k - 1$ степенями свободы.

Если $|t_j| > t_{\frac{1+\alpha}{2}}$ или $|b_j| > t_{\frac{1+\alpha}{2}} S_k \left[\sum_{i=1}^{n} P_k^2(i)\right]^{-\frac{1}{2}}$, то с доверительной вероятностью α коэффициент регрессии b_j является значимым (т. е. существенно отличным от нуля).

Основным практическим преимуществом полиномов Чебышева является то, что при повышении степени полинома необходимо вычислять только один дополнительный коэффициент регрессии без пересчета всех остальных.

Последовательность анализа нелинейной регрессии с помощью полиномов Чебышева состоит в следующем. Сначала находится линейная регрессия низшего порядка ($k = 1$) и соответствующая ей дисперсия $\tilde{S}_1^2$. Далее вычисляется еще один коэффициент регрессии b_2 и дисперсия $\tilde{S}_2^2$, соответствующая квадратичной регрессии. Затем критерием Фишера проверяется значимость уменьшения дисперсии $\tilde{S}_2^2$ по сравнению с дисперсией $\tilde{S}_1^2$. Если $\dfrac{\tilde{S}_1^2}{\tilde{S}_2^2} \geqslant F_\alpha$ $(f_1 = n - 2; f_2 = n - 3)$, то квадратичная регрессия предпочтительнее линейной.

Кубическая регрессия предпочитается квадратичной, если $\dfrac{\tilde{S}_2^2}{\tilde{S}_3^2} \geqslant F_\beta$ $(f_1 = n - 3;$ $f_2 = n - 4)$.

Таблица 237

Полиномы Чебышева $P_j(i)$ и значения $P^* = \sum_{i=1}^{n} P_k^2(i)$

n	i	$P_1(i)$	$P_2(i)$	$P_3(i)$	$P_4(i)$	n	i	$P_1(i)$	$P_2(i)$	$P_3(i)$	$P_4(i)$
3	1	−1	1			11	P^*	110	858	4290	286
	2	0	−2			12	1	−11	55	−33	33
	3	1	1				2	−9	25	3	−27
	P^*	2	6				3	−7	1	21	−33
4	1	−3	1	−1			4	−5	−17	25	−13
	2	−1	−1	3			5	−3	−29	19	12
	3	1	−1	−3			6	−1	−35	7	28
	4	3	1	1			P^*	572	12012	5148	8008
	P^*	20	4	20		13	1	−6	22	−11	99
5	1	−2	2	−1	1		2	−5	11	0	−66
	2	−1	−1	2	−4		3	−4	2	6	−96
	3	0	−2	0	6		4	−3	−5	8	−54
	4	1	−1	−2	−4		5	−2	−10	7	11
	5	2	2	1	1		6	−1	−13	4	64
6	P^*	10	14	10	70		7	0	−14	0	84
	1	−5	5	−5	1		P^*	182	2002	572	68068
	2	−3	−1	7	−3	14	1	−13	13	−143	143
	3	−1	−4	4	2		2	−11	7	−11	−77
	4	1	−4	−4	2		3	−9	2	66	−132
	5	3	−1	−7	−3		4	−7	−2	98	−92
	6	5	5	5	1		5	−5	−5	95	−13
	P^*	70	84	180	28		6	−3	−7	67	63
7	1	−3	5	−1	3		7	−1	−8	24	108
	2	−2	0	1	−7		P^*	910	728	97240	136136
	3	−1	−3	1	1	15	1	−7	91	−91	1001
	4	0	−4	0	6		2	−6	52	−13	−429
	P^*	28	84	6	154		3	−5	19	35	−869
8	1	−7	7	−7	7		4	−4	−8	58	−704
	2	−5	1	5	−13		5	−3	−29	61	−249
	3	−3	−3	7	−3		6	−2	−44	49	251
	4	−1	−5	3	9		7	0	−56	0	756
	P^*	168	168	264	616		P^*	280	37128	39780	6466460
9	1	−4	28	−14	14	16	1	−15	35	−455	273
	2	−3	7	7	−21		2	−13	21	−91	−91
	3	−2	−8	13	−11		3	−11	9	143	−221
	4	−1	−17	9	9		4	−9	−1	267	−201
	5	0	−20	0	18		5	−7	−9	301	−101
	P^*	60	2772	990	2002		6	−5	−15	265	23
10	1	−9	6	−42	18		7	−3	−19	179	129
	2	−7	2	14	−22		8	−1	−21	63	189
	3	−5	−1	35	−17		P^*	1360	5712	1007760	470288
	4	−3	−3	31	3	17	1	−8	40	−28	52
	5	−1	−4	12	18		2	−7	25	−7	−13
	P^*	330	132	8580	2860		3	−6	12	7	−39
11	1	−5	15	−30	6		4	−5	1	15	−39
	2	−4	6	6	−6		5	−4	−8	18	−24
	3	−3	−1	22	−6		6	−3	−15	17	−3
	4	−2	−6	23	−1		7	−2	−20	13	17
	5	−1	−9	14	4		8	−1	−23	7	31
	6	0	−10	0	6		9	0	−24	0	36

Окончание таблицы 237

n	i	$P_1(i)$	$P_2(i)$	$P_3(i)$	$P_4(i)$	n	i	$P_1(i)$	$P_2(i)$	$P_3(i)$	$P_4(i)$
17	P^*	408	7752	3876	16796	19	7	−3	−21	112	42
18	1	−17	68	−68	68		8	−2	−26	83	227
	2	−15	44	−20	−12		9	−1	−29	44	352
	3	−13	23	13	−47		10	0	−30	0	396
	4	−11	5	33	−51		P^*	570	13566	213180	2288132
	5	−9	−10	42	−36	20	1	−19	57	−969	1938
	6	−7	−22	42	−12		2	−17	39	−357	−102
	7	−5	−31	35	13		3	−15	23	85	−1122
	8	−3	−37	23	33		4	−13	9	377	−1402
	9	−1	−40	8	44		5	−11	−3	539	−1187
	P^*	1938	23256	23256	28424		6	−9	−13	591	−687
19	1	−9	51	−204	612		7	−7	−21	553	−77
	2	−8	34	−68	−68		8	−5	−27	445	503
	3	−7	19	28	−388		9	−3	−31	287	948
	4	−6	6	89	−453		10	−1	−33	99	1188
	5	−5	−5	120	−354		P^*	2660	17556	4903140	22881320
	6	−4	−14	126	−168						

Процесс вычислений заканчивается, как только

$$\frac{\tilde S_k^2}{\tilde S_{k+1}^2} < F_\alpha \ \ (f = n-k-1; f_2 = n-k-2)\,.$$

Так как повышение степени полинома от k до $k+1$ не уменьшает дисперсию регрессии, принимается уравнение регрессии, описываемое полиномом степени k.

Задача 375. *В результате исследований получена следующая зависимость:*

x_i: 1,1; 3,2; 5,3; 7,4; 9,5; 11,6; 13,7; 15,8; 17,9; 20,0

y_i: 1,3; 4,75; 6,8; 1,86; −15,6; −51,1; −110,3; −198,6; −321,8; −485,2.

Необходимо найти уравнение регрессии и определить порядок его нелинейности, приняв достоверность $\alpha = 0{,}95$.

Имеем $x^* = \dfrac{x-1{,}1}{2{,}1} = \dfrac{x}{2{,}1} - \dfrac{1{,}1}{2{,}1}$, тогда $x_i^* = i$. Результаты вычислений сводим в таблицу, там же приводим значения полиномов Чебышева для $n = 10$ и $k = 1, 2, 3, 4$ из табл. 237:

$x_i^* = i$	y_i	$P_1(i)$	$y_iP_1(i)$	$P_2(i)$	$y_iP_2(i)$	$P_3(i)$	$y_iP_3(i)$	$P_4(i)$	$y_iP_4(i)$
1	1,30	−9	−11,70	6	7,80	−42	−54,60	18	23,40
2	4,75	−7	−33,25	2	9,50	14	66,50	−22	−104,50
3	6,80	−5	−34,00	−1	−6,80	35	238,00	−17	−115,60
4	1,86	−3	−5,58	−3	−5,58	31	57,66	3	5,58
5	−15,60	−1	15,60	−4	62,40	12	−187,20	18	−280,80
6	−51,10	1	−51,10	−4	204,40	−12	613,20	18	−919,80
7	−110,30	3	−330,90	−3	330,90	−31	3419,30	3	−330,90
8	−198,60	5	−993,00	−1	198,60	−35	6951,00	−17	3376,20
9	−321,80	7	−2252,60	2	−643,60	−14	4505,20	−22	7079,60
10	−485,20	9	−4366,80	6	−2911,20	42	−20378,40	18	−8733,60
$\sum$	−1167,89		−8063,33		−2753,58		−4769,34		−0,42
$\sum_{i=1}^{10} P_k^2(i)$		330		132		8580		2860	

Вычисляем $b_0 = \frac{1}{10} \cdot \sum_{i=1}^{10} y_i = \frac{1}{10} \cdot (-1167{,}89) = -116{,}79$; $b_1 = \frac{\sum_{i=1}^{10} y_i \cdot P(i)}{\sum_{i=1}^{10} P_1^2(i)} = -\frac{8063{,}33}{330} = -24{,}4343$.

Уравнение регрессии имеет вид

$$y = -116{,}79 - 24{,}434 \cdot (x^* - \bar{x}^*) = 17{,}597 - 24{,}434 \cdot x^*.$$

Находим дисперсию

$$\tilde{S}_1^2 = \frac{S_1^2}{n-2}, \quad \text{где} \quad S_1^2 = S_0^2 - b_1^2 \cdot \sum_{i=1}^{n} P_1^2(i); \quad \text{где} \quad \sum_{i=1}^{n} y_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^{n} y_i\right)^2.$$

Имеем

$$S_0^2 = 393510{,}8521 - \frac{1}{10} \cdot 1167{,}89^2 = 257114{,}1469;$$

$$S_1^2 = 257114{,}1469 - 24{,}434^2 \cdot 330 = 60097{,}42942; \quad \tilde{S}_1^2 = \frac{1}{8} \cdot 60097{,}42942 = 7512{,}1787.$$

Вычислим коэффициент $b_2 = \frac{\sum_{i=1}^{10} y_i \cdot P_2(i)}{\sum_{i=1}^{10} P_2^2(i)} = -\frac{2753{,}58}{132} = -20{,}86.$

Далее находим

$$S_2^2 = 60097{,}42942 - 20{,}86^2 \cdot 132 = 2659{,}002; \quad \tilde{S}_2^2 = \frac{1}{7} \cdot 2659{,}002 = 379{,}857.$$

Из таблиц находим $F_{0,95}(f_1 = 8,\ f_2 = 7) = 3{,}726$.

Так как $\frac{\tilde{S}_1^2}{\tilde{S}_2^2} = \frac{7512{,}1787}{379{,}857} = 19{,}776 > F_{0,95}\ (8; 7) = 3{,}726$, следует признать, что дисперсия $\tilde{S}_2^2$ значимо меньше, чем $\tilde{S}_1^2$. Следовательно, квадратичная регрессия предпочтительнее линейной и на этом этапе следует принять уравнение регрессии $y = -106{,}569 - 24{,}434 \times (x^* - \bar{x}^*) - 20{,}86 \cdot (x^* - \bar{x}^*)^2$.

Выполним теперь вычисления для кубической регрессии:

$$b_3 = \frac{\sum_{i=1}^{10} y_i \cdot P_3(i)}{\sum_{i=1}^{10} P_3^2(i)} = -\frac{4769{,}34}{8580} = -0{,}5558;$$

$$S_3^2 = S_2^2 - b_3^2 \cdot \sum_{i=1}^{10} P_3^2(i) = 2659{,}002 - 0{,}5558^2 \cdot 8580 = 8{,}52297;$$

$$\tilde{S}_3^2 = \frac{1}{6} \cdot 8{,}52297 = 1{,}42049; \quad \frac{\tilde{S}_2^2}{\tilde{S}_3^2} = \frac{379{,}857}{1{,}42} = 267{,}5.$$

Отношение $\frac{\tilde{S}_2^2}{\tilde{S}_3^2}$ существенно превышает критическое значение $F_{0,95}(f_1 = 7, f_2 = 6) = 4{,}20$.

Продолжаем вычисления:

$$b_4 = \frac{\sum_{i=1}^{10} y_i \cdot P_4(i)}{\sum_{i=1}^{10} P_4^2(i)} = -\frac{0{,}420}{2860} = -1{,}468 \cdot 10^{-4};$$

$$S_4^2 = S_3^2 - \left(1{,}468 \cdot 10^{-4}\right)^2 \cdot \sum_{i=1}^{10} P_4^2(i) = 8{,}5229; \quad \tilde{S}_4^2 = \frac{1}{5} \cdot 8{,}5229 = 1{,}704.$$

Тогда $\dfrac{\tilde{S}_3^2}{\tilde{S}_4^2} = \dfrac{1{,}42}{1{,}704} < 1$, что означает, что увеличение степени регрессионного полинома свыше трех не уменьшает, а увеличивает дисперсию регрессии.

Поэтому следует принять в качестве адекватной кубическую регрессионную модель

$$y = -106{,}569 - 24{,}434 \cdot (x^* - \bar{x}^*) - 20{,}86 \cdot (x^* - \bar{x}^*)^2 - 0{,}5558 \cdot (x^* - \bar{x}^*)^3.$$

При переходе к натуральной переменной $x^* = \dfrac{x}{2{,}1} - \dfrac{1{,}1}{2{,}1}$ имеем

$$y = -106{,}569 - 24{,}434 \cdot \left(\frac{x}{2{,}1} - \frac{1{,}1}{2{,}1} - 5{,}5\right) - 20{,}86 \cdot \left(\frac{x}{2{,}1} - \frac{1{,}1}{2{,}1} - 5{,}5\right)^2 - \\ - 0{,}5558 \cdot \left(\frac{x}{2{,}1} - \frac{1{,}1}{2{,}1} - 5{,}5\right)^3 = -588{,}328 + 77{,}344 \cdot x - 2{,}36 \cdot x^2 - 0{,}0581 \cdot x^3.$$

Проверяем значимость коэффициентов регрессии.

Для b_1 имеем $f = n - 2 = 8$, $t_{0,975}(8) = 2{,}31$ (см. табл. 118) и

$$|b_1| = 24{,}434 < 2{,}31 \cdot \frac{S_1}{\sqrt{\sum\limits_{i=1}^{10} P_1^2(i)}} = 2{,}31 \cdot \frac{\sqrt{60037{,}429}}{\sqrt{330}} = 31{,}17.$$

Следовательно, коэффициент b_1 незначимо отличается от нуля и линейный член может быть исключен из уравнения регрессии. Далее для коэффициента b_2 имеем $f = n - 3 = 7$, $t_{0,975}(7) = 2{,}45$ и

$$|b_2| = 20{,}86 > 2{,}37 \cdot \frac{\sqrt{2659{,}002}}{\sqrt{132}} = 10{,}63.$$

Следовательно, коэффициент регрессии b_2 значимо отличается от нуля.

Для b_3 имеем $f = n - 4 = 6$, $t_{0,975}(6) = 2{,}45$ и

$$|b_3| = 0{,}5558 > 2{,}45 \cdot \frac{\sqrt{8{,}523}}{\sqrt{8580}} = 0{,}077.$$

Следовательно, коэффициент регрессии b_3 также значимо отличается от нуля. Итак, с учетом значимости коэффициентов регрессии, окончательное уравнение регрессии будет иметь вид

$$y = -106{,}569 - 20{,}86 \cdot \left(\frac{x}{2{,}1} - \frac{1{,}1}{2{,}1} - 5{,}5\right)^2 - 0{,}5558 \cdot \left(\frac{x}{2{,}1} - \frac{1{,}1}{2{,}1} - 5{,}5\right)^3 = \\ = 735{,}518 + 88{,}852 \cdot x - 2{,}36 \cdot x^2 - 0{,}0581 \cdot x^3.$$

5.3.4. Выбор наилучшей регрессионной модели по Вильямсу–Клуту

Ранее (см. раздел 5.3.1.3) были рассмотрены методы сравнения регрессий в случае их одинакового функционального вида и одинаковых независимых переменных. Рассмотрим теперь задачу сравнения любых двух или более регрессий независимо от их функционального вида и количества независимых переменных.

Из двух любых регрессий наилучшей признается та, которая имеет меньшую остаточную сумму квадратов (меньшую дисперсию рассеяния зависимой переменной вокруг линии регрессии).

Вильямс и Клут [14] для сравнения двух любых регрессий y_1 и y_2 без трудоемкого предварительного подсчета их остаточных дисперсий предложили критерий,

основанный на оценке с помощью обычной линейной регрессии углового коэффициента λ линии

$$y - \frac{y_1 + y_2}{2} = \lambda(y_2 - y_1),$$

где y — наблюдаемое значение зависимой переменной; y_1, y_2 — предсказываемые сравниваемыми регрессиями значения зависимой переменной.

Значимые отрицательные значения углового коэффициента λ указывают на то, что регрессия y_1 лучше, чем регрессия y_2. Значимые положительные значения λ указывают на предпочтительность регрессии y_2. Если λ незначимо отличается от нуля, то модели y_1 и y_2 признаются равноценными. Значимость отклонения λ от нуля проверяется методами обычного регрессионного анализа для простой линейной модели. В этом случае испытываемой линейной регрессионной моделью является

$$z = \lambda x, \quad \text{где} \quad z = y - \frac{1}{2}(y_1 + y_2); \quad x = y_2 - y_1.$$

Оценка углового коэффициента линейной регрессии подсчитывается по формуле (см. раздел 5.3.1.1.1)

$$\lambda = \frac{n\sum_{i=1}^{n} z_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} z_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}.$$

Далее находим

$$S^2 = \frac{1}{n-2}\sum_{i=1}^{n}(z_i - \lambda x_i)^2; \quad S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2; \quad S_\lambda = \frac{S}{S_x\sqrt{n-1}}.$$

Тогда, если:

$-\lambda < -t_\alpha S_\lambda$, то регрессия y_2 предпочитается регрессии y_1;

$\lambda > t_\alpha S_\lambda$, то регрессия y_1 предпочитается регрессии y_2;

$-t_{\frac{1+\alpha}{2}} S_\lambda < \lambda < t_{\frac{1+\alpha}{2}} S_\lambda$, то обе регрессии признаются эквивалентными.

Здесь t_α — α-квантиль распределения Стьюдента с $f = n - 2$.

Задача 376. *Выбрать критерием Вильямса–Клута при доверительной вероятности $\alpha = 0{,}95$ лучшую из двух регрессий*

$$y_2 = 1{,}61 + 3{,}99 \cdot x \quad \text{и} \quad y_1 = 1{,}24 + 2{,}81 \cdot x + 0{,}2 \cdot x^2$$

для следующего набора данных:

x_i:	2	4	6	9	11	16	17	20	25	31;
y_i:	9	19	22	41	49	61	69	83	98	128.

Результаты промежуточных расчетов приведены в таблице. Имеем $x_i^* = y_{2i} - y_{1i}$ и $z_i = y_i - \dfrac{y_{1i} + y_{2i}}{2}$;

$$\sum_{i=1}^{10} x_i^* = 378{,}845; \quad \sum_{i=1}^{10} {x_i^*}^2 = 38861{,}13893; \quad \left(\sum_{i=1}^{10} x_i^*\right)^2 = 134523{,}534;$$

$$\sum_{i=1}^{10} z_i = -199{,}55; \quad \sum_{i=1}^{10} x_i^* \cdot z_i = -19245{,}586.$$

$$\lambda = \frac{10 \cdot (-19345{,}586) - 378{,}845 \cdot (-199{,}55)}{10 \cdot 38861{,}13893 - 143523{,}534} = -0{,}497, \quad \text{т. е.} \quad z = -0{,}497 \cdot x^*.$$

x_i	y_i	y_{2i}	y_{1i}	$\frac{y_{1i}+y_{2i}}{2}$	$y_{2i}-y_{1i}$	$z_i = y_i - \frac{y_{1i}+y_{2i}}{2}$
2	9	9,59	7,66	8,625	−1,930	0,375
4	19	17,57	15,68	16,625	−1,890	2,375
6	22	25,55	25,30	26,425	−0,125	−3,425
9	41	37,52	42,73	40,125	5,210	0,875
11	49	45,50	56,35	50,925	10,850	−1,925
16	61	65,45	97,40	81,425	31,950	−20,425
17	69	69,44	106,81	88,125	37,370	−19,125
20	83	81,41	137,44	109,425	56,030	−26,425
25	98	101,36	196,49	148,925	95,130	−50,925
30	128	125,30	280,55	202,925	155,250	−74,925
				$\sum$	378,845	−193,55

Для $\alpha = 0{,}95$ и $f = n-2 = 10-2 = 8$ из табл. 118 имеем $t_{0,95}(8) = 1{,}86$. Далее вычисляем

$$S_{x^*}^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i^* - \bar{x}^*)^2 = 2647 \quad (S_{x^*} = 51{,}45)\,;$$

$$S^2 = \frac{1}{n-2}\cdot\sum_{i=1}^{10}(z_i - \lambda\cdot x_i^*)^2 = 10{,}272 \quad (S = 3{,}205)\,; \qquad S_\lambda = \frac{3{,}205}{51{,}45\cdot\sqrt{10-1}} = 0{,}027.$$

Имеем $\lambda = -0{,}497 < -t_{0,95}\cdot S_\lambda = -1{,}86\cdot 0{,}0207 = -0{,}039$. Следовательно, линейная регрессия y_2 предпочтительнее квадратичной y_1.

5.3.5. Прогнозирование по регрессии

Имеется регрессия $y = f(x)$, оцененная по равноотстоящим значениям x_i $(i = 1, 2, \ldots, n)$. Требуется найти значение зависимой переменной y_{n+l} при $x = x_{n+l}$ (l — глубина экстраполяции) и оценить ошибку прогноза. Двусторонний α-доверительный интервал для y_{n+l} имеет вид [600]

$$f(x_{n+l}) - t_{\frac{1+\alpha}{2}} s\lambda(n,l) = y_{n+l}^{\text{н}} \leqslant y_{n+l} \leqslant y_{n+l}^{\text{в}} = f(x_{n+l}) + t_{\frac{1+\alpha}{2}} s\lambda(n,l)\,,$$

где $f(x_{n+l}) = y_{n+l}$ — значение y, вычисленное по регрессии при $x = x_{n+l}$; t_α — α-квантиль распределения Стьюдента при $f = n-k-1$ степенях свободы; $S^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}[y_i - f(x_i)]^2$ — дисперсия регрессии; k — порядок регрессии; λ — коэффициент, зависящий от n и l.

Коэффициент $\lambda(n,l)$ равен:

— для линейной регрессии

$$\lambda(n,l) = \left\{\frac{(n+1)\left(n^2-1\right) + 3\,(n+2l-1)^2}{n\,(n-1)}\right\}^{\frac{1}{2}};$$

— для квадратичной регрессии

$$\lambda(n,l) = \left\{1 + \frac{(n+l)^2}{\sum\limits_{i=1}^{n} i^2} + \frac{\sum\limits_{i=1}^{n} i^4 - 2\,(n+1)^2\sum\limits_{i=1}^{n} i^2 + n\,(n+l)^4}{n\sum\limits_{i=1}^{n} i^4 - \left(\sum\limits_{i=1}^{n} i^2\right)^2}\right\}^{\frac{1}{2}}.$$

Для удобства вычислений напомним, что

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}; \quad \sum_{i=1}^{n} i^4 = \frac{n(n+1)(2n+1)\left(3n^2+3n-1\right)}{30}.$$

Задача 377. *По данным эксперимента:*

x_i: 1 2 3 4 5 6 7 8 9 10;
y_i: 7 9 10 6 6 −4 −10 −22 −33 −47

найти 95%-й доверительный интервал для $y = f(x)$ при $x = 15$.

Учитывая, что в нашем случае $x_i = i$, применим непосредственно технику полиномов Чебышева (см. раздел 5.3.3.2). Результаты вычислений (с применением табл. 237 для $P_k(i)$) приведены в таблице:

$x_i = i$	y_i	$P_1(i)$	$y_iP_1(i)$	$P_2(i)$	$y_iP_2(i)$	$P_3(i)$	$y_iP_3(i)$	$P_4(i)$	$y_iP_4(i)$
1	7	−9	−63	6	63	−42	−294	18	126
2	9	−7	−63	2	18	14	126	−22	−198
3	10	−5	−50	−1	−10	35	350	−17	−170
4	6	−3	−18	−3	−18	31	186	3	18
5	2	−1	−2	−4	−8	12	24	18	36
6	−4	1	−4	−4	16	−12	48	18	−72
7	−10	3	−30	−3	30	−31	310	3	−30
8	−22	5	−110	−1	22	−35	770	−17	374
9	−33	7	−231	2	−66	−14	462	−22	726
10	−47	9	−423	6	−282	42	−1974	18	−846
$\sum$	−82		−994		−235		8		−36
$\sum_{i=1}^{n} P_k^2(i)$		330		132		8580		2860	

Вычисляем $b_0 = -\frac{82}{10} = -8{,}2$; $b_1 = \frac{\sum_{i=1}^{10} y_i \cdot P_1(i)}{\sum_{i=1}^{10} P_1^2(i)} = -\frac{994}{330} = -3{,}01.$

Для линейной регрессии ($\bar{x} = 5{,}5 \to y = -8{,}2 - 3{,}01 \cdot (x - 5{,}5)$) находим

$$S_0^2 = \sum_{i=1}^{10} y_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^{10} y_i\right)^2 = 4168 - 672{,}4 = 3495{,}6;$$

$$S_1^2 = S_0^2 - b_1^2 \cdot \sum_{i=1}^{10} P_1^2(i) = 3495{,}6 - 3{,}01^2 \cdot 330 = 505{,}767; \quad \tilde{S}_1^2 = \frac{S_1^2}{n-2} = \frac{505{,}767}{8} = 63{,}22.$$

Вычисляем дальше:

$$b_2 = \frac{\sum_{i=1}^{10} y_2 \cdot P_2(i)}{\sum_{i=1}^{10} P_2^2(i)} = -\frac{235}{132} = -1{,}78;$$

$$S_2^2 = S_1^2 - b_2^2 \cdot \sum_{i=1}^{10} P_2^2(i) = 505{,}767 - 1{,}78^2 \cdot 132 = 87{,}538; \quad \tilde{S}_2^2 = \frac{S_2^2}{n-3} = \frac{87{,}538}{7} = 12{,}505.$$

Критическое значение $F_{0{,}95}(f_1 = 8; f_2 = 7) = 3$.

Так как $\dfrac{\tilde{S}_1^2}{\tilde{S}_2^2} = \dfrac{63{,}22}{12{,}505} = 5{,}055 > F_{0,95}(8;7) = 3{,}7$, следует признать, что квадратичная регрессия предпочтительнее линейной.

Далее находим

$$b_{36} = \frac{\sum\limits_{i=1}^{10} y_{i=14,588} \cdot P_3(i)}{\sum\limits_{i=1}^{10} P_3^2(i)} = \frac{8}{8580} = 9{,}32 \cdot 10^{-4};$$

$$S_3^2 = S_2^2 - b_3^2 \cdot \sum_{i=1}^{10} P_3^2(i) = 87{,}538 - \left(9{,}32 \cdot 10^{-4}\right)^2 \cdot 8580 = 87{,}530; \quad \tilde{S}_3^2 = \frac{S_3^2}{6} = 14{,}588.$$

Так как $\dfrac{\tilde{S}_2^2}{\tilde{S}_3^2} = \dfrac{12{,}505}{14{,}588} < 1$, следует признать, что квадратичная регрессия предпочтительнее линейной.

Далее находим

$$y = b_0 + b_1 \cdot (x - \bar{x}) + b_2 \cdot (x - \bar{x})^2 =$$
$$= -8{,}2 - 3{,}01 \cdot (x - 5{,}5) - 1{,}78 \cdot (x - 5{,}5)^2 = -45{,}48 + 16{,}57 - 1{,}78 \cdot x^2.$$

Из уравнения регрессии для $x_{10+5} = x_{15} = 15$ вычисляем $y_{x=15} = -197{,}44$.

Ранее мы имели $\tilde{S}_2^2 = 15{,}318$ ($\tilde{S}_2 = 3{,}913$). Далее вычисляем:

$$\sum_{i=1}^{10} i^2 = \frac{10 \cdot 11 \cdot 21}{6} = 385; \quad \sum_{i=1}^{10} i^4 = \frac{10 \cdot 11 \cdot 21 \cdot (300 + 30 - 1)}{30} = 25333;$$

$$\lambda(10{,}5) = \left\{1 + \frac{225}{385} + \frac{25333 - 450 \cdot 385 + 506250}{10 \cdot 25333 - 148225}\right\}^{\frac{1}{2}} = 2{,}234.$$

Из табл. 118 находим $t_{\frac{1+0,95}{2}}$ $(f = n - k - 1)$; $t_{0,975}(7) = 2{,}37$ (здесь $k = 2$ — степень полинома). Окончательно получаем искомый доверительный интервал:

$$-197{,}44 - 2{,}37 \cdot 3{,}913 \cdot 2{,}234 = -218{,}158 \leqslant y_{x=15} \leqslant -176{,}722 = -197{,}44 + 2{,}37 \cdot 3{,}913 \cdot 2{,}234.$$

5.3.6. Специальные методы сглаживания экспериментальных данных

Рассмотренные в предыдущих разделах оценки коэффициентов регрессии методом наименьших квадратов хороши только тогда, когда остатки распределены нормально. Если же это не так, если „хвосты“ распределения „тяжелы“ или в выборке присутствуют грубые выбросы, то эти оценки быстро теряют свою эффективность.

В таких ситуациях предпочтительнее использовать иные специальные методы устойчивой (робастной) регрессии.

Для робастной регрессии оценки уравнения регрессии являются решениями задачи [620] $\sum \varphi\left(y_i - a_0 - \sum\limits_{j=1}^{n} a_j x_{ij}\right) = \min$, где $\varphi(\varepsilon)$ — функция потерь.

Очевидно, что $\varphi(\varepsilon) = \varepsilon^2$ соответствует обыкновенному методу наименьших квадратов. В некоторых ситуациях более эффективными могут быть иные функции потерь:

$\varphi(\varepsilon) = |\varepsilon|$ — метод наименьших модулей;

$\varphi(\varepsilon) = |\varepsilon|^\nu$, $1 < \nu < 2$ — итеративный метод Форсайта [605];

$$\varphi(\varepsilon) = \begin{cases} 1 - \cos\left(\dfrac{\varepsilon}{c}\right), & \text{если } |\varepsilon| < c\pi; \\ 0, & \text{если } |\varepsilon| > c\pi \ \ (c = 1{,}5 \text{ или } 2,1) \end{cases} \text{— метод Андрюса.}$$

По скорости вычислений и простоте использования метод наименьших квадратов — наилучший, однако по точности оценивания регрессии такой вывод сделать нельзя [527, 603]. Сравнение различных методов оценки регрессии приведено в [601]. Ниже мы рассмотрим некоторые из робастных методов оценки регрессии и другие специальные приемы сглаживания экспериментальных данных.

5.3.6.1. Метод наименьших модулей

В отличие от метода наименьших квадратов, при котором оценка коэффициентов регрессии ищется из условия минимизации суммы квадратов отклонений экспериментальных точек от выборочной регрессии, в [604, 606, 607, 609] развит метод наименьших модулей таких отклонений (МНМ). В соответствии с МНМ оценки коэффициентов регрессии ищутся из условия минимизации суммы абсолютных величин отклонений экспериментальных точек от выборочной регрессии.

Для линейной регрессии $y = \alpha + \beta x$ метод наименьших квадратов соответствует условию $\sum (y_i - \alpha - \beta x_i)^2 = \min$, а МНМ условию $\delta = \sum |y_i - \alpha - \beta x_i| = \min$.

Замена квадратов отклонений абсолютными величинами в ряде случаев может оказаться предпочтительнее. Прежде всего, использование МНМ позволяет ослабить влияние больших отклонений (в методе наименьших квадратов они возводятся в квадрат). В случае метода наименьших квадратов прямая регрессии имеет вид

$$\hat{y} = \bar{y} + \frac{\overline{xy} - \bar{x}\bar{y}}{S_x^2}\,(x - \bar{x}).$$

В случае же МНМ уравнение линейной регрессии есть

$$\hat{y} = \frac{x_k y_l - x_l y_k}{x_k - x_l} + \frac{y_k - y_l}{x_k - x_l} x,$$

где (x_k, y_k) и (x_l, y_l) — некоторые две выбранные из условия $\delta = \min$ экспериментальные точки.

Задача отыскания минимума является типичной задачей кусочно-выпуклого линейного программирования. В настоящей работе мы не будем излагать основы линейного программирования, а любопытного читателя отправим к специальной литературе, например, [610]. Здесь мы опишем только один простейший алгоритм решения задачи для случая линейной регрессии, рассмотренный в [606]. Для последовательности пар $(x_1, y_1), \ldots, (x_n, y_n)$ сначала предполагаем, что регрессия проходит через последнюю точку (x_n, y_n). Будем искать к ней пару — точку (x_l, y_l), минимизирующую сумму абсолютных величин отклонений δ. Для этого вычислим $(n-1)$ угловых коэффициентов

$$\beta_1 = \frac{y_n - y_1}{x_n - x_1}; \quad \beta_2 = \frac{y_n - y_2}{x_n - x_2}; \quad \ldots; \quad \beta_{n-1} = \frac{y_n - y_{n-1}}{x_n - x_{n-1}}$$

и упорядочим их по величине (например, $\beta_1 < \beta_2 < \ldots < \beta_{n-1}$).

Затем в соответствии с упорядоченным рядом β_i выстраиваем знакопеременный ряд $\pm(x_n - x_1), \pm(x_n - x_2), \ldots, \pm(x_n - x_{n-1})$ и ищем минимум (по абсолютной величине) суммы величин $\pm(x_n - x_i)$, $i = 1, \ldots, n-1$, с одной переменой знака в последовательности, выбирая соответствующим образом знаки у суммируемых величин (у первой величины всегда выбираем $+$). Фиксируем номера величин β_i и β_{i+1}, между которыми происходит перемена знака у суммы знаменателей пары (x_l, y_l). Если сумма величин $\pm(x_n - x_i)$ отрицательна, то в качестве пары к последней точке (x_n, y_n) берется точка (x_l, y_l), соответствующая угловому коэффициенту β_{i+1}, если сумма положительна, то точка, соответствующая β_i. Если сумма равна нулю, то берется любая точка $\beta_i \leqslant \beta \leqslant \beta_{i+1}$.

На втором этапе считаем, что прямая регрессии проходит через точку (x_l, y_l), и в соответствии с описанным алгоритмом ищем пару к ней. Перебор продолжается до тех пор, пока не будет получена уже найденная однажды точка. Тогда она и предпоследняя найденная являются теми двумя точками, через которые проходит искомая прямая регрессии. Аналитическое описание алгоритма весьма усложнено, поэтому дадим пояснение его реализации на количественном примере. Напомним пользователю, что задачи в предлагаемой книге — это не только иллюстрации к изложенному, но и способ пояснения методов и приемов, изложение которых иным способом слишком громоздко.

Следует отметить, что методы статистического анализа уравнений МНМ-регрессий отсутствуют. Если для способа наименьших квадратов изучены распределения параметров регрессии и их основные статистические свойства, что, собственно, и позволяет выполнить необходимый статистический анализ, то для МНМ эти распределения неизвестны. Поэтому ничего нельзя сказать о статистических свойствах регрессии, полученной МНМ-методом. Это является существенным недостатком МНМ-метода, ограничивающим его применение в статистической практике.

Задача 378. *Для набора данных:*

x_i:	1	2	3	4	5;
y_i:	38	24	31	39	48

найти прямую регрессии методом наименьших модулей.

Выбираем в качестве начальной точки последнюю точку с координатами ($x = 5$, $y = 48$). Вычисляем угловые коэффициенты

$$\beta_1 = \frac{48-38}{5-1} = 2{,}5; \quad \beta_2 = \frac{48-24}{5-2} = 8; \quad \beta_3 = \frac{48-31}{5-3} = 8{,}5; \quad \beta_4 = \frac{48-39}{5-4} = 9.$$

Ранжируя полученные значения β_i по возрастанию, получаем знакопеременный ряд для знаменателей $\pm 4, \pm 3, \pm 2, \pm 1$. Его минимальная (по абсолютной величине) сумма равна $+4-3-2-1=-2$. Перемена знака происходит между β_1 и β_2, а так как сумма отрицательна, то выбираем β_2, т. е. в пару к последней точке $(5, 48)$ берем пару, соответствующую β_2 — это точка $(2, 24)$.

Далее, предполагая, что регрессия проходит через точку $(2, 24)$, подбираем пару к ней. Имеем

$$\beta_1 = \frac{38-24}{1-2} = -14; \quad \beta_2 = \frac{31-24}{3-2} = 7; \quad \beta_3 = \frac{39-24}{4-2} = 7{,}5; \quad \beta_4 = \frac{48-24}{5-2} = 8.$$

Ранжированием по возрастанию полученного ряда β_i приходим к знакопеременному ряду знаменателей $\pm 1, \pm 1, \pm 2, \pm 3$, минимальная сумма которого равна $+1+1+2-3 = +1$. Перемена знака в сумме происходит между β_3 и β_4.

Так как сумма положительна, выбираем β_3 и соответствующую точку с координатами $x = 4$ и $y = 39$. Для этой точки имеем ряд β_i

$$\beta_1 = \frac{39-38}{4-1} = \frac{1}{3}; \quad \beta_2 = \frac{39-24}{4-2} = 7{,}5; \quad \beta_3 = \frac{39-31}{4-3} = 8; \quad \beta_4 = \frac{39-48}{4-5} = 9.$$

Для ранжированного по возрастанию ряда β_i получаем знакопеременный ряд ± 3, ± 2, ± 1, ± 1, для которого наименьшая сумма равна $+3-2-1-1=-1$.

Перемена знака происходит между β_1 и β_2, и, так как сумма отрицательна, то выбираем β_2 и точку с координатами $(2, 24)$, но эта точка уже встречалась. Следовательно, искомая прямая проходит через две точки — предыдущую $(4, 39)$ и полученную повторением $(2, 24)$. Отсюда имеем $x_k = 4$, $y_k = 39$ и $x_l = 2$, $y_l = 24$. Тогда уравнение прямой регрессии имеет вид

$$y = \frac{4 \cdot 24 - 2 \cdot 39}{4-2} + \frac{39-24}{4-2} \cdot x = 9 + 7{,}5 \cdot x,$$

Для нее сумма модулей отклонений от экспериментальных точек равна

$$\sum_{i=1}^{5}|y_i - 9 - 7{,}5 \cdot x| = |38 - 9 - 7{,}5| + |24 - 9 - 7{,}5 \cdot 2| + \ldots + |48 - 9 - 7{,}5 \cdot 5| = 23{,}5$$

Для примера: прямая регрессии, полученная методом наименьших квадратов, была бы $y = 25{,}5 + 3{,}5 \cdot x$, и для нее

$$\sum_{i=1}^{5}|y_i - 25{,}5 - 3{,}5x_i| = |38 - 25{,}5 - 3{,}5| + \ldots + |48 - 25{,}5 - 17{,}5| = 36.$$

Суммы квадратов отклонений равны

$$\sum_{i=1}^{5}(y_i - 9 - 7{,}5 \cdot x_i)^2 = 464{,}75, \quad \text{и} \quad \sum_{i=1}^{5}(y_i - 25{,}5 - 3{,}5 \cdot x_i)^2 = 311{,}5,$$

что соответствует назначению обеих оценок регрессии.

5.3.6.2. Метод последней точки

В ряде задач, в основном связанных с прогнозами в экономике, для линии регрессии важно получить наибольшее приближение к последним точкам, что позволяет использовать ее с большей эффективностью для прогноза последующего поведения изучаемого процесса.

В [606] рассмотрен метод последней точки, когда уравнение прямой регрессии принудительно приводится в последнюю точку (x_n, y_n) за счет некоторого увеличения остаточной дисперсии регрессии. При этом надежда возлагается на то, что увеличение остаточной дисперсии будет скомпенсировано повышением точности прогноза в области последней точки. Уравнение регрессии для метода последней точки имеет вид [606]

$$y = y_n + \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i - n y_n \sum_{i=1}^{n} x_i - n x_n \sum_{i=1}^{n} y_i + n^2 x_n y_n}{\left(\sum_{i=1}^{n} x_i\right)^2 - 2 n x_n \sum_{i=1}^{n} x_i + n^2 x_n^2} (x - x_n).$$

Следует отметить, что метод последней точки может быть эффективно использован совместно с методом наименьших модулей (см. раздел 5.3.6.1). Тогда, выбрав последнюю точку (x_n, y_n) в качестве начальной, методами МНМ к ней ищут оптимальную пару.

Задача 379. *Для данных задачи* 378 *найти прямую регрессии методом последней точки.*

Имеем $x_n = 5$; $y_n = 48$; $\sum_{i=1}^{n} x_i = 15$; $\sum_{i=1}^{5} y_i = 180$; $\left(\sum_{i=1}^{5} x_i\right)^2 = 225$.

Получаем

$$y = 44 + \frac{32 \cdot 150 - 5 \cdot 44 \cdot 32 - 5 \cdot 12 \cdot 150 + 25 \cdot 12 \cdot 44}{1024 - 2 \cdot 5 \cdot 12 \cdot 32 + 25 \cdot 144} \cdot (x - 12) = 14 + 2{,}5 \cdot x.$$

5.3.6.3. Метод однозначной аппроксимации

Рассмотренные ранее методы регрессионного анализа на основе принципа наименьших квадратов не позволяют получить обратимые уравнения регрессии, т. е. совокупность экспериментальных точек (x, y) не аппроксимируется однозначно.

Регрессии y по x и x по y для одного и того же набора экспериментальных данных не совпадают.

В [608] рассмотрен метод однозначной аппроксимации, основанный на минимизации суммы площадей:

$$S = \sum_{i=1}^{n} S_i^2 = \sum_{i=1}^{n} \left[y_i - f(x_i)\right]\left[x_i - f^{-1}(y_i)\right] = \min,$$

где x_i, y_i — координаты i-й экспериментальной точки; $f(x)$ — искомая функция; $f^{-1}(x)$ — функция, обратная к функции $f(x)$.

Для модели $y = \alpha + \beta x$ оценки коэффициентов α и β находятся из условия $\sum\limits_{i=1}^{n}(y_i - \varepsilon - \beta x_i)\left(x_i - \dfrac{y_i - \alpha}{\beta}\right) = \min.$

В соответствии с [608] оценки α и β, удовлетворяющие сформулированным требованиям, равны соответственно

$$b = \pm\left\{\frac{n\sum\limits_{i=1}^{n} y_i^2 - \left(\sum\limits_{i=1}^{n} y_i\right)^2}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2}\right\}^{\frac{1}{2}}; \quad a = \bar{y} - b\bar{x}$$

(знак + или − выбирается исходя из знака корреляции между переменными y и x).

При оценке коэффициентов по методу однозначной аппроксимации регрессии y по x и x по y совпадают.

В [602] рассмотрена модификация метода однозначной аппроксимации, предполагающая минимизацию суммы квадратов расстояний от экспериментальных точек до искомой прямой (ортогональная регрессия). Для случая ортогональной регрессии оценки равны

$$b = \frac{2\left(n\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i\right)}{n^3\left(S_x^2 - S_y^2\right) + \sqrt{n^6 S_x S_y\left(S_x^2 - S_y^2\right) + 4\left(n\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i\right)^2}}; \quad a = \bar{y} - b\bar{x}.$$

Коэффициенты уравнения ортогональной регрессии зависят от масштаба переменных, в то время как для обычной регрессии и метода однозначной аппроксимации такая зависимость отсутствует. Корректное обобщение метода однозначной аппроксимации (и ортогональной регрессии) приведено в работе [638], содержащей исследование уравнения регрессии в форме нормального уравнения прямой.

Задача 380. *Для набора данных* ($n = 10$):

x_i:	2	4	5	7	6	9	14	21	32	29;
y_i:	8	15	14	16	18	25	31	50	65	61

найти оценки коэффициентов регресии методом однозначной аппроксимации.

Имеем

$$\sum_{i=1}^{10} x_i = 303; \quad \bar{x} = \frac{1}{n}\cdot\sum_{i=1}^{10} x_i = 30{,}3; \quad \left(\sum_{i=1}^{10} x_i\right)^2 = 91809; \quad \sum_{i=1}^{10} x_i^2 = 13097;$$

$$\sum_{i=1}^{10} y_i = 129; \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{10} y_i = 12{,}9; \quad \left(\sum_{i=1}^{10} y_i\right)^2 = 16641; \quad \sum_{i=1}^{10} y_i^2 = 2713;$$

$$\sum_{i=1}^{10} x_i \cdot y_i = 5924; \quad S_x^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{10} (x_i - \bar{x})^2 = 435{,}122 \quad (S_x = 20{,}86);$$

$$S_y^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{10} (y_i - \bar{y})^2 = 116{,}54 \quad (S_y = 10{,}79);$$

Вычисляем оценки методом однозначной аппроксимации (выбираем знак +, так как корреляция между y и x положительна)

$$b = \left\{\frac{10 \cdot 2713 - 16641}{10 \cdot 13097 - 91809}\right\}^{\frac{1}{2}} = 0{,}517; \quad a = \bar{y} - b\bar{x} = 12{,}9 - 0{,}517 \cdot 30{,}3 = -2{,}765.$$

Следовательно, получаем уравнение регрессии $y = -2{,}765 + 0{,}517 \cdot x$. Из него можно непосредственно получить уравнение регрессии x по y: $x = 5{,}348 + 1{,}934 \cdot y$.

Оценки уравнения регрессии y по x, полученные методом наименьших квадратов (см. раздел 5.3.1.1.1), привели бы к уравнению $y = -2{,}701 + 0{,}515x$.

Легко видеть, что оценки методом однозначной аппроксимации достаточно близки к оценкам наименьших квадратов.

5.3.6.4. Метод обратных разделенных разностей

Метод предложен в [637]. Его авторы предлагают искать математическую модель регрессии в классе цепных дробей

$$y = \frac{\sum_{i=0}^{n} a_i x^i}{\sum_{i=0}^{n} b_i x^i}.$$

Однако приведенные авторами работы [637] численные эксперименты не позволяют сделать вывод о предпочтительности этого метода по сравнению с классическим методом наименьших квадратов. По крайней мере, возможный выигрыш в частных задачах никоим образом не компенсируется резко возрастающей сложностью расчетов.

5.3.6.5. Метод условно-относительных разностей

Завершая настоящий раздел достаточно „экзотичных" методов получения эмпирических зависимостей, дополним его кратким указанием на еще один метод, предложенный недавно — метод условно-относительных разностей [635, 636]. Этот метод предполагает минимизацию не суммы квадратов вида $\sum [y_i - f(x_i)]^2 = \min$, а квадратов относительных разностей

$$\sum \frac{[y_i - f(x_i)]^2}{y_i^2} = \min.$$

Естественно, что применение метода условно-относительных разностей обеспечивает повышение точности аппроксимации в области малых значений y_i, но снижает ее в области больших значений y_i.

Имея некоторые преимущества при решении частных задач, этот метод, в силу вычислительной сложности и неясности статистических свойств оценок такого вида, вряд ли может быть рекомендован для широкого применения.

5.4. Контрольные карты

Этот раздел обращен прежде всего к инженерам и специалистам производства, для которых контроль качества создаваемой ими продукции — один из важнейших заключительных этапов длительного цикла изготовления изделия.

Совокупность контролируемых показателей качества изделия представляется, если отвлечься от их физической природы, потоком случайных величин, в непрерывно меняющемся хаосе которого контролер должен своевременно рассмотреть тревожные симптомы ухудшения качества изделия или признаки нарушения технологического процесса его изготовления.

Ознакомившись с предыдущими разделами книги, читатель справедливо может заключить, что мы рассмотрели столько всевозможных методов сравнения случайных совокупностей, обнаружения тренда и всего другого, что этого должно быть вполне достаточно для решения любых задач, в том числе и задач контроля качества производства. Он будет по сути прав.

Контрольные карты — это прежде всего самая распространенная форма применения статистики в производстве. Сегодня статистические карты контроля качества стали настолько общепринятыми, что любая достаточно большая фирма, не пользующаяся ими в том или ином виде, оказывается в невыгодном положении по сравнению с конкурентами.

Идея контрольной карты состоит в следующем. Делается предположение о распределении совокупности или выборочных статистик совокупности. Отбираются независимые случайные повторные выборки из текущей продукции. Подсчитываются и наносятся на специально подготовленную контрольную карту выборочные значения контролируемых статистик вместе с дву- или односторонними (в зависимости от специфики контроля) доверительными интервалами. Если текущие измерения показателей качества продукции находятся внутри доверительного интервала — процесс производства статистически управляем, качество продукции находится в допустимых пределах; если нет — требуется регулирующее вмешательство в производственный процесс.

По сути метод контрольных карт — это многократное повторение во времени метода доверительных интервалов, но в отличие от него он более динамичен.

Динамичность контрольных карт дает возможность фиксировать, когда и где наблюдаемые величины выходят за рамки случайных колебаний, указывая на необходимость регулирования технологического процесса.

Техника контрольных карт, организация их внедрения достаточно широко освещены в литературе [14, 22, 97, 611]. Мы приведем краткие математические и табличные описания наиболее распространенных методов, которые могут служить кратким введением для инженера или специалиста, не знакомившегося ранее с ними, и позволят применять их в своей практике.

5.4.1. Контрольные карты Шухарта

Первые контрольные карты предложены Шухартом [612] в 1931 г. Он предложил состав таких карт, технологию их составления, методы расчета и правила пользования.

Контрольные карты можно рассматривать как многократное применение критерия значимости. Суть контрольных карт состоит в следующем. На карту наносятся контрольные границы параметра, соответствующие выбранной достоверности контроля. Затем на карту наносятся выборочные значения контролируемого параметра. Если его положение выходит за контрольные границы — это сигнал к необходимости вмешательства в процесс.

5.4.1.1. $\bar{x}$- и R-карты

Предположим, что в последовательно контролируемых выборках объема n измеряется некоторая характеристика изделия — x. В каждой выборке вычисляется среднее значение контролируемого признака $\bar{x}$ и его размах $R = x_{\max} - x_{\min}$. Обозначим их для i-й выборки через $\bar{x}_i$ и R_i. В $\bar{x}$-карте контрольные границы для среднего при достоверности $\alpha = 0{,}997$ будут

$$x_{\text{н}} = \bar{\bar{x}} - A\bar{R} \text{ (нижняя граница)}; \quad x_{\text{в}} = \bar{\bar{x}} + A\bar{R} \text{ (верхняя граница)},$$

где $\bar{\bar{x}} = \dfrac{1}{k}\sum \bar{x}_i$ — среднее по k текущим выборкам; $\bar{R} = \dfrac{1}{k}\sum R_i$ — средний размах по k текущим выборкам; A — коэффициент, значения которого приведены в табл. 238.

Таблица 238

Коэффициенты для вычисления границ регулирования контрольных карт Шухарта [1]

Объем выборки, n	$\bar{x}$-карта	R-карта		S-карта		
	A	D_1	D_2	B_1	B_2	C
2	1,880	0	3,267	0	2,298	0,798
3	1,023	0	2,575	0	2,111	0,886
4	0,729	0	2,282	0	1,982	0,921
5	0,577	0	2,115	0	1,889	0,940
6	0,483	0	2,004	0,085	1,817	0,951
7	0,419	0,076	1,924	0,158	1,762	0,960
8	0,373	0,136	1,864	0,215	1,715	0,965
9	0,337	0,184	1,816	0,262	1,676	0,969
10	0,308	0,223	1,777	0,302	1,644	0,973
11	0,285	0,256	1,744	0,336	1,616	0,976
12	0,266	0,284	1,716	0,365	1,589	0,977
13	0,249	0,308	1,692	0,392	1,568	0,980
14	0,235	0,329	1,671	0,414	1,548	0,981
15	0,223	0,348	1,652	0,434	1,530	0,982
16	0,212	0,364	1,636	0,454	1,514	0,984
17	0,203	0,379	1,621	0,469	1,499	0,984
18	0,194	0,392	1,608	0,486	1,486	0,986
19	0,187	0,404	1,596	0,500	1,472	0,986
20	0,180	0,414	1,586	0,513	1,461	0,987
21	0,173	0,425	1,575	0,525	1,451	0,988
22	0,167	0,434	1,566	0,536	1,440	0,988
23	0,162	0,443	1,557	0,546	1,432	0,989
24	0,157	0,452	1,548	0,556	1,422	0,989
25	0,153	0,459	1,541	0,566	1,414	0,990

Для R-карты средним уровнем (центральным) является $\bar{R}$, нижние и верхние границы регулирования равны соответственно

$$R_{\text{н}} = D_1\bar{R} \quad \text{и} \quad R_{\text{в}} = D_2\bar{R},$$

где D_1 и D_2 — коэффициенты, приведенные в табл. 238.

Задача 381. *В результате контроля 25 выборок изделий, при объеме каждой выборки $n = 5$, получены следующие значения контролируемого параметра x:*

Номер выборки i	Выборочные значения x_{ij}					Статистики			
	x_1	x_2	x_3	x_4	x_5	$\bar{x}_i$	R_i	S_i^2	S_i
1	1	−1	−2	2	0	0,0	4	2,5	1,58
2	−2	0	−2	2	0	−0,4	4	2,8	1,67
3	0	0	1	−1	3	0,6	4	2,3	1,52
4	1	2	−1	2	−2	0,4	4	3,3	1,82
5	2	1	−1	1	0	0,6	3	1,3	1,14
6	2	−1	2	−1	0	0,4	3	2,3	1,52
7	−2	−1	0	−2	0	−1,0	2	1,0	1,00
8	1	1	0	1	−1	0,4	2	0,8	0,89
9	0	−2	1	−1	4	0,4	6	5,3	2,30
10	−1	−1	0	−2	1	−0,6	3	1,3	1,14
11	1	−1	2	−1	0	0,2	3	1,7	1,30
12	2	2	−1	−1	0	0,4	3	2,3	1,52
13	0	−2	2	−2	0	−0,4	4	2,8	1,67
14	0	−1	1	−1	2	0,2	3	1,7	1,30
15	−3	0	0	0	−1	−0,8	3	1,7	1,30
16	0	−1	0	1	2	0,4	3	1,3	1,14
17	3	1	−3	−2	0	−0,2	6	5,7	2,39
18	−1	−2	−1	−1	0	−1,0	2	0,5	0,71
19	0	0	0	1	1	0,4	1	0,3	0,55
20	0	−5	2	2	0	−0,2	7	8,2	2,86
21	2	−1	−1	0	0	0,4	3	1,3	1,14
22	0	0	−3	2	−3	−0,8	5	4,7	2,17
23	3	1	4	−2	−1	1,0	6	6,5	2,55
24	1	−3	0	1	−1	−0,4	4	2,8	1,67
25	−1	0	1	0	1	0,2	2	0,4	0,63

Вычислить границы для контрольных $\bar{x}$- и R-карт Шухарта ($k = 25$).

Имеем $\bar{\bar{x}} = \frac{1}{25} \cdot \sum\limits_{i=1}^{25} \bar{x}_i = \frac{0{,}2}{25} = 0{,}008$; $\bar{R} = \frac{1}{k} \cdot \sum\limits_{i=1}^{k} R_i = \frac{90}{25} = 3{,}6$.

Контрольные границы для среднего (для $n = 5$ из табл. 238 находим $A = 0{,}577$) будут равны

$$x_н = \bar{\bar{x}} - A \cdot \bar{R} = 0{,}008 - 0{,}577 \cdot 3{,}60 = -2{,}069;$$

$$x_в = \bar{\bar{x}} + A \cdot \bar{R} = 0{,}008 + 0{,}577 \cdot 3{,}60 = 2{,}085.$$

Мы видим, что нигде $\bar{x}_i$ не вышли за контрольные границы.

Теперь для R-карты из табл. 238 имеем $D_1 = 0$ и $D_2 = 2{,}115$.

Тогда $R_н = 0$ и $R_в = 2{,}115 \cdot 3{,}60 = 7{,}61$. Видим, что ни одно значение R_i не выходит за эти пределы, т. е. процесс статистически управляем.

$\bar{x}$- и R-карты Шухарта рекомендуется применять при $n < 10 \div 20$.

5.4.1.2. s-карта

s-карта для выборочного контроля среднего квадратичного отклонения более чувствительна к изменению рассеяния, чем R-карта. Алгоритм построения и использования s-карты состоит в следующем. По текущим k выборкам вычисляется оценка дисперсии контролируемого признака

$$s^2 = \frac{\sum\limits_{i=1}^{k} \nu_i S_i^2}{\sum\limits_{i=1}^{k} \nu_i}, \quad \text{где} \quad \nu_i = n_i - 1; \quad s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 .$$

Если объемы выборок n_i одинаковы, то $s^2 = \frac{1}{k}\sum_{i=1}^{k} s_i^2$. Центральная линия s-карты равна $s_ц = Cs$, а границы зоны регулирования равны $s_н = B_1 s$ и $s_в = B_2 s$, где B_1, B_2 и C — коэффициенты, приведенные в табл. 238.

Задача 382. *Для данных задачи* 381 *построить контрольные границы S-карты Шухарта.*

Имеем $s^2 = \frac{1}{25}\cdot\sum_{i=1}^{25} s_i^2 = 2{,}592$ $(s = 1{,}610)$.

Следовательно, для контрольных границ имеем $s_ц = 0{,}940\cdot 1{,}610 = 1{,}513$; $s_н = 0\cdot 1{,}610 = 0$; $s_в = 1{,}889\cdot 1{,}610 = 3{,}041$ (предварительно из табл. 238 для $n = 5$ получили $C = 0{,}940$, $B_1 = 0$ и $B_2 = 1{,}889$).

Видим, что из всех выборок только s_i в 20-й выборке близко к $s_в = 3{,}04$ (впрочем, то же самое было и для R-карты).

5.4.1.3. $\bar{x}$- и s-карты для выборок неравного объема

Пусть имеем k выборок объемами n_i $(i = 1, 2, \ldots, k)$. Если объемы выборок не равны между собой, то, естественно, границы регулирования $\bar{x}$- и s-карт будут меняться от выборки к выборке. В этом случае следует поступать следующим образом. По всем выборкам находим

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} n_i\bar{x}_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}}{\sum_{i=1}^{k} n_i} \quad \text{и} \quad s^2 = \frac{\sum_{i=1}^{k}(n_i - 1)\, s_i^2}{\sum_{i=1}^{k}(n_i - 1)}.$$

Для $\bar{x}$-карты имеем

$$x_ц = \bar{\bar{x}} \text{ — центральная линия;} \quad x_н = \bar{\bar{x}} - A_i s \text{ — нижняя граница;}$$
$$x_в = \bar{\bar{x}} + A_i s \text{ — верхняя граница;}$$

для s-карты имеем

$$s_ц = C_i s \text{ — центральная линия;} \quad s_н = B_{1i} s \text{ — нижняя граница;}$$
$$s_в = B_{2i} s \text{ — верхняя граница.}$$

Здесь C_i, B_{1i}, B_{2i} — коэффициенты из табл. 238, определяемые для объема каждой выборки; A_i — коэффициенты, приведенные в табл. 239.

Таблица 239

Значения коэффициентов A_i для $\bar{x}$-карт Шухарта при неравных объемах контролируемых выборок

n_i	A_i	n_i	A_i	n_i	A_i
2	2,121	10	0,949	18	0,770
3	1,732	11	0,905	19	0,688
4	1,500	12	0,866	20	0,671
5	1,342	13	0,832	21	0,655
6	1,225	14	0,802	22	0,640
7	1,134	15	0,775	23	0,626
8	1,061	16	0,750	24	0,612
9	1,000	17	0,728	25	0,600

Задача 383. *В результате контроля получены результаты измерений параметра x (средние значения и дисперсии, полученные в результате контроля в выборках разного объема, приведены в таблице). Необходимо найти контрольные границы для $\bar{x}$- и s-карт.*

Номер выборки, i	Объем выборки, n_i	$\bar{x}$-карта					s-карта		
		$\bar{x}_i$	s_i^2	s_i	$x_{\text{н}i}$	$x_{\text{в}i}$	$s_{\text{ц}i}$	$s_{\text{н}i}$	$s_{\text{в}i}$
1	8	1356	1281	35,8	1353	1440	39,55	8,18	70,30
2	10	1380	2570	50,7	1357,6	1435,4	39,88	12,38	67,39
3	12	1448	1901	43,6	1361	1432	40,05	14,96	65,13
4	7	1358	1384	37,2	1350	1443	40,57	16,07	64,27
5	10	1372	1430	37,8	1357,6	1435,4	40,21	16,97	63,45
6	6	1430	1672	40,9	1346,3	1446,7	40,25	17,79	62,71
7	8	1356	1354	36,8	1353	1440	40,33	18,61	62,06
8	6	1426	643	25,3	1346,3	1446,7	40,33	19,22	61,44
9	5	1444	3418	58,5	1341,5	1451,5	40,42	19,92	60,91
10	12	1404	1608	40,1	1361	1432	40,42	20,49	60,34

Для $\bar{x}$-карты имеем среднюю линию

$$x_{\text{ц}} = \frac{\sum n_i \cdot x_i}{\sum n_i} = \frac{8 \cdot 1356 + 10 \cdot 1380 + \ldots + 12 \cdot 1404}{8 + 10 + \ldots + 12} = 1396{,}5.$$

Коэффициенты A_i определяются для каждой выборки по табл. 239. Например, для первой выборки имеем $n_1 = 8$ и $A_1 = 1{,}061$, для второй выборки $n_2 = 10$ и $A_2 = 0{,}949$, и т. д.

Находим оценку средней дисперсии

$$s^2 = \frac{\sum (n_i - 1) \cdot s_i^2}{\sum (n_i - 1)} = 1680{,}4 \quad (S = 40{,}99).$$

Тогда для первой выборки имеем

$$x_{\text{н}1} = \bar{\bar{x}} - A_1 \cdot s = 1396{,}5 - 1{,}061 \cdot 40{,}99 = 1353; \quad x_{\text{в}1} = \bar{\bar{x}} + A_1 \cdot s = 1396{,}5 + 1{,}061 \cdot 40{,}99 = 1440.$$

Полученные по аналогии результаты для других объемов выборок приведены в таблице.

Для расчета средней линии и контрольных границ s-карты пользуемся коэффициентами C_i, B_{1i} и B_{2i} из табл. 238, взятыми для соответствующих объемов выборок n_i.

Например, для первой выборки имеем $n_1 = 8$, $C_1 = 0{,}965$, $B_{11} = 0{,}215$ и $B_{21} = 1{,}75$. Следовательно, средняя линия равна

$$s_{\text{ц}1} = C_1 \cdot s = 0{,}965 \cdot 40{,}99 = 39{,}55,$$

а границы регулирования равны:

$$s_{\text{н}1} = B_{11} \cdot s = 0{,}215 \cdot 40{,}99 = 8{,}18; \quad s_{\text{в}1} = B_{21} \cdot s = 1{,}715 \cdot 40{,}99 = 70{,}30.$$

Вычисленные по аналогии значения $s_{\text{ц}i}$, $s_{\text{н}i}$ и $s_{\text{в}i}$ для остальных выборок приведены в таблице.

Из анализа видно, что среднее значение в третьей выборке выходит за границы регулирования, а все стандартные отклонения находятся в пределах границ регулирования.

5.4.1.4. Контрольная карта для доли дефектных изделий (p-карта)

Предположим, наблюдается k выборок объемами n_i, причем в i-й выборке фиксируется m_i дефектных изделий ($1 \leqslant i \leqslant k$). Оценкой доли дефектных изделий в i-й выборке является величина $p_i = \dfrac{m_i}{n_i}$.

Оценка среднего значения доли дефектных изделий $\bar{p} = \dfrac{\sum\limits_{i=1}^{k} m_i}{\sum\limits_{i=1}^{k} n_i}$ является центральной линией p-карты. Границы регулирования определяются формулами

$$p_{\text{н}i} = \bar{p} - 3\left\{\frac{\bar{p}(1-\bar{p})}{n_i}\right\}^{\frac{1}{2}}; \quad p_{\text{в}i} = \bar{p} + 3\left\{\frac{\bar{p}(1-\bar{p})}{n_i}\right\}^{\frac{1}{2}}.$$

Если все значения n_i отличаются друг от друга незначительно, то используется единая пара контрольных границ

$$p_{\text{нг}} = \bar{p} - 3\left\{\frac{\bar{p}}{\bar{n}}(1-\bar{p})\right\}^{\frac{1}{2}}; \quad p_{\text{вг}} = \bar{p} + 3\left\{\frac{\bar{p}}{\bar{n}}(1-\bar{p})\right\}^{\frac{1}{2}}, \quad \text{где} \quad \bar{n} = \frac{1}{k}\sum_{i=1}^{k} n_i.$$

Задача 384. *В результате контроля различных партий получены результаты, приведенные в таблице. Найти границы регулирования p-карты.*

Номер партии, i	Объем выборки, n_i	Число дефектов, m_i	Доля дефектных изделий, p_i	Границы регулирования	
				$p_{\text{н}i}$	$p_{\text{в}i}$
1	981	27	0,028	0,0234	0,0622
2	1422	87	0,061	0,0449	0,0771
3	1174	87	0,074	0,0563	0,0917
4	1524	76	0,050	0,0344	0,0655
5	1353	80	0,059	0,0425	0,0755
6	847	25	0,030	0,0091	0,0509
7	1535	37	0,024	0,0085	0,0395
8	1248	19	0,015	0,0022	0,0322
9	1296	49	0,038	0,0211	0,0549
10	985	42	0,043	0,0236	0,0623

Имеем оценку средней доли дефектных изделий

$$\bar{p} = \frac{\sum\limits_{i=1}^{10} m_i}{\sum\limits_{i=1}^{10} n_i} = \frac{27 + 87 + 87 + \ldots + 42}{981 + 1422 + \ldots + 985} = 0{,}0428.$$

Далее, для каждой выборки вычисляется граница регулирования $p_{\text{н}i}$ и $p_{\text{в}i}$. Например, для $i = 1$ имеем $n_1 = 981$ и

$$p_{\text{н}1} = 0{,}0428 - 3 \cdot \left\{\frac{0{,}0428 \cdot (1 - 0{,}0428)}{981}\right\}^{\frac{1}{2}} = 0{,}0234;$$

$$p_{\text{в}2} = 0{,}0428 + 3 \cdot \left\{\frac{0{,}0428 \cdot (1 - 0{,}0428)}{981}\right\}^{\frac{1}{2}} = 0{,}0622.$$

Рассчитанные по аналогии величины для остальных выборок приведены в таблице, из которой видно, что доли дефектных изделий во всех выборках находятся в пределах границ регулирования.

5.4.1.5. Контрольная граница числа дефектов (c-карта)

В серии последовательных выборок фиксируются количества дефектных изделий c_i. Контрольная карта для регулирования процесса по числу дефектных изделий включает в себя:

— центральную линию $c_ц = \bar{c} = \frac{1}{k}\sum_{i=1}^{k} c_i$;

— нижнюю границу регулирования $c_н = \bar{c} - 3\sqrt{\bar{c}}$;

— верхнюю границу регулирования $c_в = \bar{c} + 3\sqrt{\bar{c}}$.

Рекомендуется при объеме выборок $n \geqslant 30$.

Задача 385. *Для данных задачи 384 вычислить параметры c-карты.*

Имеем $\bar{c} = \frac{1}{10}\cdot\sum c_i = 52{,}9$; $c_н = 52{,}9-3\cdot\sqrt{52{,}9} = 31{,}08$; $c_в = 52{,}9+3\cdot\sqrt{52{,}9} = 74{,}72$. Видим, что процесс находится в контролируемом состоянии.

5.4.1.6. Карты индивидуальных значений и скользящего размаха

Карты этого типа используются для контроля индивидуальных значений параметра x и их колеблемости.

Пусть x_i — i-е текущее значение контролируемого параметра, а $R_i = |x_{i-1} - x_i|$ $(i > 1)$ — абсолютная величина разностей последовательных пар значений x. Контроль значений x_i и R_i выполняется с помощью карты скользящего размаха.

Абсолютные значения разностей $|x_{i-1} - x_i|$ можно рассматривать как размахи последовательных выборок объема $n = 2$.

Центральной линией карты x_i является величина

$$x_ц = \bar{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i;$$

ее нижние и верхние границы равны соответственно

$$x_н = \bar{\bar{x}} - 2{,}59675\bar{R}; \quad x_в = \bar{\bar{x}} + 2{,}59675\bar{R}.$$

Для $\bar{R}$ имеем центральную линию $R_ц = \bar{R} = \frac{1}{n-1}\sum_{i=1}^{n} R_i$, нижнюю границу $R_н = 0$ и верхнюю границу $R_в = 3{,}267\bar{R}$.

Карты индивидуальных значений (x_i-карты) и скользящего размаха (R_i-карты) наиболее удобны в тех случаях, когда наблюдения берутся через большие промежутки времени. Особенно это важно для изделий, выпускаемых штучно, из которых формирование контролируемых выборок затруднено.

Задача 386. *В результате контроля параметра x_i десяти изделий получены результаты, приведенные в таблице. Найти параметры x_i- и R_i-карт и проверить по ним стабильность параметра изделия.*

Номер изделия, i	Значение параметра, x_i	Скользящий размах $R_i = \lvert x_{i-1} - x_i \rvert$	Номер изделия, i	Значение параметра, x_i	Скользящий размах $R_i = \lvert x_{i-1} - x_i \rvert$
1	20,1		6	15,1	2,0
2	19,4	0,7	7	21,2	6,1
3	18,2	1,2	8	24,3	3,1
4	27,4	4,2	9	25,1	0,8
5	17,1	5,3	10	20,0	5,1

Находим

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{10} \cdot 202{,}9 = 20{,}29; \quad \bar{R} = \frac{1}{9} \cdot \sum R_i = \frac{28{,}5}{9} = 3{,}167.$$

Для x_i-карты имеем

$$x_{ц} = \bar{\bar{x}} = 20{,}29; \quad x_{н} = \bar{\bar{x}} - 2{,}6596 \cdot \bar{R} = 20{,}29 - 2{,}6596 \cdot 3{,}167 = 11{,}867;$$
$$x_{в} = \bar{\bar{x}} + 2{,}6596 \cdot \bar{R} = 20{,}29 + 2{,}6596 \cdot 3{,}167 = 28{,}172.$$

Видим, что все индивидуальные значения x_i находятся в поле регулируемого состояния.

Для R_i-карты получаем (из таблицы имеем $R_{ц} = \bar{R} = 3{,}167$): $R_{н} = 0$ и $R_{в} = 3{,}267 \times 3{,}167 = 10{,}346$. Ни одно из текущих значений R_i не выходит за пределы зоны регулируемого состояния.

5.4.2. Контрольные карты накопленных сумм (ККНС)

Контрольные карты накопленных сумм получили широкое применение в последние 30 лет. Карты этого типа более чувствительны, чем карты Шухарта, к скачкообразным изменениям параметров. Теория и практика применения ККНС рассмотрены в [613–615].

Техника ККНС по существу является реализацией последовательной выборочной процедуры. Если в рассмотренных выше картах Шухарта на них наносятся точки, соответствующие отдельным наблюдениям, то точки, наносимые на ККНС, не соответствуют отдельным наблюдениям, вычисленным по одной выборке. Все они, начиная с исходного, дают информацию о наблюдениях от первого до последнего. В ККНС рассматриваемая точка равняется значению текущего наблюдения плюс значение некоторой статистики, вычисленной по предшествующей выборке.

Рассмотрим алгоритм формирования ККНС и принятия решения с их помощью. Через y_i будем обозначать значение контролируемого параметра в i-й выборке, а через $\tilde{y}_i^{н(в)}$ — нижние (верхние) контрольные границы, с которыми сравнивается полученное значение y_i.

В отличие от карт Шухарта контролируемый параметр y_i является суммой всех предыдущих наблюдений, а контрольные границы $\tilde{y}_i$ меняются с каждым вновь полученным наблюдением. Контрольные границы являются линейной функцией каждого наблюдения, т. е.

$$\tilde{y}_m^{н(в)} = a_m \mp bi, \quad i = 1, \ldots, m, \quad \text{где} \quad a_m = y_m \pm b(m+d).$$

Таким образом, для каждой новой m-й выборки находятся параметры a_m и b контрольных границ, с которыми сравниваются все предыдущие значения y_i $(i = 1, \ldots, m-1)$ контролируемого параметра.

Если контролируется нижнее значение параметра (т. е. требуется, чтобы $x_n \geqslant \mu$, где μ — нормативное значение параметра), то используется контрольная граница

$$\tilde{y}_m^{в} = a_m - bi = y_m + b\,(m+d) - bi$$

(x_n — некоторый контролируемый параметр, например, среднее или дисперсия).

Если при любом $i = 1, \ldots, m-1$ будет наблюдаться неравенство $y_i > \tilde{y}_m^{в}$, то в точке m следует признать значимый сдвиг вниз контролируемого параметра. По аналогии для контроля верхнего значения (т. е. требуется, чтобы $x_n \leqslant \mu$) используется неравенство $y_i < \tilde{y}_m^{н} = y_m - b\,(m+d) + bi$ (тогда в m-й выборке фиксируется значимый сдвиг контролируемого параметра вверх).

В случае двустороннего контроля (т. е. когда необходимо фиксировать $x_n = \mu$) используется интервальная оценка $\tilde{y}_m^{в} < y_i < \tilde{y}_m^{н}$.

Если неравенство удовлетворяется, то равенство $x_n = \mu$ не отклоняется, если нет — фиксируется значимый сдвиг параметра либо вверх, либо вниз. Параметр b в уравнении для $\tilde{y}_m^{\text{н(в)}}$ определяется минимальным сдвигом параметра, который должен быть выявлен в процессе контроля, и масштабным коэффициентом k, равным отношению числа единиц ординаты y_i, приходящихся на единицу абсциссы x_ι.

Параметр d определяется выбранным значением ошибки первого рода α (вероятностью отвергнуть правильную гипотезу), а также разбросом значений контролируемого параметра x_n (при этом предполагается, что ошибка второго рода $\beta < 0{,}01$, т. е. мощность процедуры $(1-\beta) > 0{,}99$). Подробнее технику ККНС продемонстрируем на примерах конкретных контрольных карт.

5.4.2.1. ККНС для среднего значения

Пусть x_ι — среднее i-й выборки объема n. Величиной, подлежащей контролю, является накопленная сумма

$$y_i = \frac{1}{s_{\bar{x}}} \sum_i (\bar{x}_i - \mu),$$

где μ — нормативное значение контролируемого параметра и $s_{\bar{x}}$ — стандартное отклонение среднего. Обычно рекомендуется выполнить оценку $s_{\bar{x}}$ по первым m выборкам

$$s_{\bar{x}} = \left(\frac{1}{m}\sum_{i=1}^{m} s_i^2\right)^{\frac{1}{2}}, \quad \text{где} \quad s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2,$$

n — объем выборки (предполагается, что все выборки равного объема n). Если $mn(n-1) > 30 \div 40$, то оценка достаточно точна.

Величина $D = \Delta\mu$ определяется минимальным недопустимым сдвигом среднего относительно нормативного значения μ. Параметры b и d равны: $b = \dfrac{D}{2}$, $d = -\dfrac{2}{\delta^2}\ln\alpha$, где $\delta = \dfrac{D}{s_{\bar{x}}}$, α — ошибка первого рода.

Если производится односторонний контроль (т. е. контроль того факта, что $\bar{x} > \mu$ или $\bar{x} < \mu$), то используется значение α. Если контролируется уход среднего в любую сторону (т. е. контроль того, что $\bar{x} \neq \mu$), то вместо α используется величина 2α.

Таким образом, контрольные границы, полученные после появления m-й выборки равны

— для одностороннего контроля

$$\tilde{y}_m^{\text{н}} = y_m + \frac{D}{2}\left(m - \frac{2}{\delta^2}\ln\alpha\right) - \frac{D}{2}i, \quad i = 1, 2, \ldots, m-1;$$

$$\tilde{y}_m^{\text{в}} = y_m - \frac{D}{2}\left(m - \frac{2}{\delta^2}\ln\alpha\right) + \frac{D}{2}i, \quad i = 1, 2, \ldots, m-1;$$

— для двустороннего контроля

$$\tilde{y}_m^{\text{н(в)}} = y_m \pm \frac{D}{2}\left(m - \frac{2}{\delta^2}\ln\alpha\right) \mp \frac{D}{2}i.$$

Если для некоторого значения y_i имеет место одно из неравенств $y_\iota > \tilde{y}_m^{\text{в}}$; $y_i < \tilde{y}_m^{\text{н}}$; $\tilde{y}_m^{\text{н}} < y_i < \tilde{y}_m^{\text{в}}$, то в выборке с номером m признается соответственно сдвиг среднего значения вниз, вверх или в любую сторону с вероятностью $1-\alpha$.

Легко видеть, что сдвиг в каждой ν-й выборке контролируется вычислением по полученному значению y_ν новых контрольных границ $\tilde{y}_\nu^{\text{н(в)}}$ и сравнением их с ранее полученными значениями y_i, $i < \nu$. Легче всего, естественно, использовать

специальные шаблоны [14, 97], которые позволяют избежать вычислений. Мы же поясним технику применения ККНС для среднего на конкретном примере.

Задача 387. *По данным, приведенным в таблице, найти параметры для ККНС среднего и проверить стабильность среднего по отношению к нормативному значению* $\mu = 27$, *полагая, что недопустимым является отклонение* $D = \Delta\mu = 5$ *единиц. Принять* $\alpha = 0{,}1$ *и* $n = 5$.

Номер выборки, i	$\bar{x}_i$	s_i^2	y_i	Номер выборки, i	$\bar{x}_i$	s_i^2	y_i
1	25	2,4	−1,04	6	34	3,2	7,30
2	36	6,1	3,65	7	16	2,1	1,56
3	21	4,2	0,52	8	28	4,1	2,08
4	9	3,8	1,56	9	30	4,2	3,65
5	31	2,9	2,65	10	24	3,8	2,08

Имеем $m = 10$ и находим $s_{\bar{x}}^2 = \dfrac{1}{10} \cdot \sum S_i^2 = 3{,}68$ $(s_{\bar{x}} = 1{,}918)$.

Имеем $b = D/2 = 5/2 = 2{,}5$. Далее рассмотрим задачу одностороннего контроля нижнего значения $\bar{x}_i$ (т. е. контролем должен быть выявлен факт снижения $\bar{x}$ по сравнению с $\mu = 27$ более, чем на 5 единиц).

Имеем $\delta = \dfrac{D}{s_{\bar{x}}} = 2{,}607$; $d = -\dfrac{2}{\delta^2} \cdot \ln\alpha = -\dfrac{2}{2{,}607^2} \cdot \ln 0{,}1 = 0{,}677$.

Тогда для контроля некоторой j-й ($j = 2, 3, \ldots$) выборки имеем порождаемые ею верхние контрольные границы для $i = 1, \ldots, j-1$ предыдущих значений y_i:

$$\tilde{y}_j^{\text{в}} = y_j + 2{,}5 \cdot (j + 0{,}677) - 2{,}5 \cdot i; \quad i = 1, \ldots, j-1.$$

В нашем случае они равны

$j = 2$ $(i = 1)$: $\tilde{y}_2^{\text{в}i} = 3{,}65 + 2{,}5 \cdot (2 + 0{,}677) - 2{,}5 \cdot i = 10{,}34 - 2{,}5 \cdot i$;

$j = 3$ $(i = 1{,}2)$: $\tilde{y}_3^{\text{в}i} = 0{,}52 + 2{,}5 \cdot (3 + 0{,}677) - 2{,}5 \cdot i = 9{,}71 - 2{,}5 \cdot i$;

$j = 4$ $(i = 1{,}2{,}3)$: $\tilde{y}_4^{\text{в}i} = 1{,}56 + 2{,}5 \cdot (4 + 0{,}677) - 2{,}5 \cdot i = 13{,}25 - 2{,}5 \cdot i$;

$j = 5$ $(i = 1, \ldots, 4)$: $\tilde{y}_5^{\text{в}i} = 3{,}5 + 2{,}5 \cdot (5 + 0{,}677) - 2{,}5 \cdot i = 17{,}84 - 2{,}5 \cdot i$;

$j = 6$ $(i = 1, \ldots, 5)$: $\tilde{y}_6^{\text{в}i} = 7{,}30 + 2{,}5 \cdot (6 + 0{,}677) - 2{,}5 \cdot i = 16{,}9 - 2{,}5 \cdot i$;

$j = 7$ $(i = 1, \ldots, 6)$: $\tilde{y}_7^{\text{в}i} = 1{,}56 + 2{,}5 \cdot (7 + 0{,}677) - 2{,}5 \cdot i = 29{,}75 - 2{,}5 \cdot i$;

$j = 8$ $(i = 1, \ldots, 7)$: $\tilde{y}_8^{\text{в}i} = 2{,}08 + 2{,}5 \cdot (8 + 0{,}677) - 2{,}5 \cdot i = 23{,}77 - 2{,}5 \cdot i$;

$j = 9$ $(i = 1, \ldots, 8)$: $\tilde{y}_9^{\text{в}i} = 3{,}05 + 2{,}5 \cdot (9 + 0{,}677) - 2{,}5 \cdot i = 27{,}84 - 2{,}5 \cdot i$;

$j = 10$ $(i = 1, \ldots, 9)$: $\tilde{y}_{10}^{\text{в}i} = 2{,}08 + 2{,}5 \cdot (10 + 0{,}677) - 2{,}5 \cdot i = 28{,}77 - 2{,}5 \cdot i$.

Порождаемые этими уравнениями переменные верхние контрольные границы $\tilde{y}_j^{\text{в}i}$ приведены в таблице:

j	Экспериментальные значения y_i								
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
	−1,04	3,65	0,52	1,56	3,65	7,30	1,56	2,08	3,65
2	7,84								
3	7,21	4,71							
4	10,75	8,25	5,75						
5	15,34	12,84	10,34	7,84					
6	14,19	11,69	9,19	6,69	4,19				
7	18,25	15,75	13,25	10,75	8,25	**5,75**			
8	21,27	18,77	16,27	13,77	11,27	8,77	6,27		
9	25,34	22,84	20,34	17,84	15,34	12,84	10,34	7,84	
10	26,27	23,77	21,27	18,77	16,27	13,77	11,27	8,77	6,27

Из таблицы видно, что значимый сдвиг среднего вниз произошел в седьмой выборке, так как $y_6 = 7{,}3 > \tilde{y}_7^{\text{в}6} = 5{,}75$ (выпадающее наблюдение в таблице выделено). Из примера следует одно основное преимущество ККНС — она фиксирует сдвиг среднего вниз по двум последовательным значениям контролируемых сумм, что очень важно при непрерывном контроле производства и позволяет непрерывно и оперативно на него воздействовать. Уменьшение среднего в седьмой выборке до 16 уверенно выявляется контролем с помощью ККНС.

Теперь по аналогии (для демонстрации техники и развития навыков вычисления и анализа ККНС) рассмотрим задачу сдвига среднего вверх. В этом случае необходимо использовать контрольные границы

$$\tilde{y}_j^{\text{н}i} = y_j - 2{,}5 \cdot (j + 0{,}677) + 2{,}5 \cdot i, \quad i < j.$$

В нашем случае это приводит к контрольным границам

$j = 2\ (i = 1)$: $\tilde{y}_2^{\text{н}i} = 3{,}65 - 2{,}5 \cdot 2{,}677 + 2{,}5 \cdot i = -3{,}04 + 2{,}5 \cdot i$;
$j = 3\ (i = 1, 2)$: $\tilde{y}_3^{\text{н}i} = 0{,}52 - 2{,}5 \cdot 3{,}677 + 2{,}5 \cdot i = -8{,}67 + 2{,}5 \cdot i$;
$j = 4\ (i = 1, 2, 3)$: $\tilde{y}_4^{\text{н}i} = 1{,}56 - 2{,}5 \cdot 4{,}677 + 2{,}5 \cdot i = -10{,}13 + 2{,}5 \cdot i$;
$j = 5\ (i = 1, \ldots, 4)$: $\tilde{y}_5^{\text{н}i} = 3{,}65 - 2{,}5 \cdot 5{,}677 + 2{,}5 \cdot i = -10{,}54 + 2{,}5 \cdot i$;
$j = 6\ (i = 1, \ldots, 5)$: $\tilde{y}_6^{\text{н}i} = 7{,}30 - 2{,}5 \cdot 6{,}677 + 2{,}5 \cdot i = -9{,}39 + 2{,}5 \cdot i$;
$j = 7\ (i = 1, 2, \ldots, 6)$: $\tilde{y}_7^{\text{н}i} = 1{,}56 - 2{,}5 \cdot 7{,}677 + 2{,}5 \cdot i = -17{,}63 + 2{,}5 \cdot i$;
$j = 8\ (i = 1, \ldots, 7)$: $\tilde{y}_8^{\text{н}i} = 2{,}08 - 2{,}5 \cdot 8{,}677 + 2{,}5 \cdot i = -19{,}61 + 2{,}5 \cdot i$;
$j = 9\ (i = 1, \ldots, 8)$: $\tilde{y}_9^{\text{н}i} = 3{,}65 - 2{,}5 \cdot 9{,}677 + 2{,}5 \cdot i = -20{,}54 + 2{,}5 \cdot i$;
$j = 10\ (i = 1, \ldots, 9)$: $\tilde{y}_{10}^{\text{н}i} = 2{,}08 - 2{,}5 \cdot 10{,}677 + 2{,}5 \cdot i = -24{,}61 + 2{,}5 \cdot i$.

Вычисленные контрольные границы $\tilde{y}_j^{\text{н}i}$ приведены в таблице:

j	Экспериментальные значения y_i								
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
	−1,04	3,65	0,52	1,56	3,65	7,30	1,56	2,08	3,65
2	**−0,54**								
3	−6,17	−3,67							
4	−7,63	−5,13	−2,63						
5	−8,04	−5,54	−3,04	−0,54					
6	−6,89	−4,39	−1,89	−0,61	−3,11				
7	−15,13	−12,63	−10,13	−7,63	−5,13	−2,63			
8	−17,11	−14,61	−12,11	−9,61	−7,11	−4,61	−2,11		
9	−18,04	−15,54	−13,04	−10,54	−8,04	−5,54	−3,04	−0,54	
10	−22,11	−19,61	−17,11	−14,61	−12,11	−9,61	−7,11	−4,61	−2,11

Из таблицы видно, что сдвиг среднего значения вверх происходит во второй выборке ($y_1 = -1{,}4 < \tilde{y}_1^{\text{н}1} = -0{,}54$). Таким образом, увеличение среднего с 25 в первой выборке до 36 во второй выборке является значимым, что и выявила контрольная карта.

5.4.2.2. ККНС выборочных размахов

Карты этого типа используются для контроля изменчивости рассеяния в выборках. Контролируемым параметром является сумма размахов

$$y_i = \sum_i R_i, \quad \text{где} \quad R_i = x_{i\max} - x_{i\min} \text{ — размах } i\text{-й выборки.}$$

Параметры уравнения для контрольных границ вычисляются по формулам

$$b = \sigma_0 c\nu \frac{\ln \dfrac{\sigma_1}{\sigma_0}}{1 - \dfrac{\sigma_0}{\sigma_1}}; \quad d = -2\frac{\ln \alpha}{\nu \ln \dfrac{\sigma_1}{\sigma_0}}.$$

Здесь σ_0 — нормативное значение стандартного нормального отклонения (определяется предварительно по специально проведенным испытаниям); $\frac{\sigma_1}{\sigma_0}$ — изменение значения стандартного нормального отклонения, которое должно быть выявлено с достоверностью не менее $1-\alpha$; α — ошибка первого рода; c, ν — коэффициенты, зависящие от n (приведены в табл. 240).

Таблица 240

Коэффициенты c и ν для ККНС выборочного размаха [97, 617]

n	3	4	5	6	7	8	9	10
c	0,233	0,188	0,16	0,142	0,128	0,118	0,11	0,103
ν	7,27	10,95	14,49	17,86	21,08	24,11	27,01	29,82

Практически важно контролировать увеличение разброса значений, чтобы вовремя принять меры для стабилизации контролируемого параметра. Поэтому будем рассматривать односторонние ККНС для выявления сдвига вверх значения σ. Уравнение для контрольной границы в этом случае имеет вид

$$\tilde{y}_j^{\text{н}} = y_j - b\,(j+d) + bi, \quad i < j,$$

или для нашего случая

$$\tilde{y}_j^{\text{н}} = y_j - \sigma_0 c\nu \frac{\ln\frac{\sigma_1}{\sigma_0}}{1-\frac{\sigma_0}{\sigma_1}}\left(j - 2\frac{\ln\alpha}{\nu\ln\frac{\sigma_1}{\sigma_0}}\right) + \sigma_0 c\nu \frac{\ln\frac{\sigma_1}{\sigma_0}}{1-\frac{\sigma_0}{\sigma_1}} i.$$

Если $y_i < \tilde{y}_j^{\text{н}}$ для какого-либо $i < j$, то в j-й выборке происходит значимое увеличение разброса значений контролируемой величины.

Задача 388. *В результате испытаний $m = 10$ выборок изделий объемом $n = 5$ каждая получены следующие результаты:*

i	1	2	3	4	5	6	7	8	9	10
R_i	1,4	1,6	1,5	1,8	0,9	3,1	1,5	1,6	1,8	1,4
$y_i = \sum R_i$	1,4	3,0	4,5	6,3	7,2	10,3	11,8	13,4	15,2	16,6

Предварительно известно, что $\sigma_0 = 0,5$ и критическое значение $\sigma_1 = 0,72$. Выбираем $\alpha = 0,1$. Необходимо найти параметры ККНС для размахов и проанализировать с ее помощью полученные данные.

Выбираем $k = 2$ и для $n = 5$ находим из табл. 240 $c = 0,160$ и $\nu = 14,49$.

Далее вычисляем

$$b = \frac{\sigma_0\cdot c\cdot\nu}{k}\cdot\frac{\ln\frac{\sigma_1}{\sigma_0}}{1-\frac{\sigma_0}{\sigma_1}} = 0,5\cdot 0,160\cdot 14,49\cdot\frac{\ln\frac{0,72}{0,50}}{1-\frac{0,50}{0,72}} = 1,384; \quad d = -\frac{2\cdot\ln 0,1}{14,49\cdot\ln\frac{0,72}{0,50}} = 0,871.$$

Тогда уравнения для контрольных границ имеют вид

$$\tilde{y}_j^{\text{н}} = y_j - 1,384\cdot(j+0,871) + 1,384\cdot i.$$

Результаты расчета граничных значений $\tilde{y}_j^{\text{н}}$ приведены в таблице:

j	Экспериментальные значения y_i								
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
	1,4	3,0	4,5	6,3	7,2	10,3	11,8	13,4	15,2
2	0,41								
3	0,53	1,91							
4	−0,86	0,53	1,91						
5	0,46	1,84	3,22	4,61					
6	**2,17**	**3,56**	**4,94**	**6,32**	**7,71**				
7	2,29	3,67	5,06	6,44	7,83	9,21			
8	2,50	3,88	5,27	6,65	8,04	9,42	10,80		
9	2,92	4,30	5,69	7,07	8,46	9,84	11,22	12,61	
10	2,94	4,32	5,70	7,09	8,47	9,85	11,24	12,61	14,00

Из нее видно, что разброс существенно возрастает в шестой выборке, что фиксируется переходом значений y_i в критическую зону (отмечено жирным курсивом в таблице).

5.4.2.3. ККНС для выборочных дисперсий

Рассматриваем, как и в предыдущем разделе, одностороннюю контрольную карту, выявляющую увеличение (сдвиг вверх) дисперсии. Контролируемым параметром является теперь $y_i = \sum\limits_i f_i s_i^2$, где $f_i = n_i - 1$, n_i — объем i-й выборки. На оси абсцисс откладывается величина $x_i = \sum\limits_i f_i$. Параметры для уравнения контрольной границы находятся по формулам [97], где σ_0^2, σ_1^2 — как и выше (см. раздел 5.4.2.2), соответственно нормативное (σ_0^2) и допустимое (σ_1^2) значения дисперсии.

Уравнение для контрольной границы имеет вид

$$\tilde{y}_j^{\text{н}i} = y_j - b\,(x_j + d) + bx_i, \quad i < j.$$

В нашем случае $\tilde{y}_j^{\text{н}i} = y_j - \dfrac{2\sigma_0^2 \ln \dfrac{\sigma_1}{\sigma_0}}{1 - \left(\dfrac{\sigma_0}{\sigma_1}\right)^2} \left(x_j - \dfrac{\ln \alpha}{\ln \dfrac{\sigma_1}{\sigma_0}} \right) + \dfrac{2\sigma_0^2 \ln \dfrac{\sigma_1}{\sigma_0}}{1 - \left(\dfrac{\sigma_0}{\sigma_1}\right)^2} i.$

Если $y_i < \tilde{y}_j^{\text{н}i}$ $(i < j)$, то сдвиг дисперсии вверх (рост разброса контролируемых величин) признается значимым.

Задача 389. *Для данных, приведенных в таблице, найти параметры ККНС для дисперсий и провести анализ данных при* $\alpha = 0{,}1$:

Номер выборки, i	n_i	$\sum(n_i - 1)$	s_i^2	$\sum(n_i - 1)s_i^2$
1	5	4	3,12	12,48
2	10	13	4,10	49,38
3	8	20	3,90	76,68
4	9	28	5,10	117,48
5	8	33	9,20	163,48
6	5	37	4,80	182,68
7	5	41	4,10	199,08
8	7	47	3,80	224,16
9	6	52	4,50	246,66
10	7	58	5,00	276,66

Известно, что $\sigma_0^2 = 2$ и $\sigma_1^2 = 2{,}5$.

Вычисляем параметры уравнения контрольных границ:

$$b = \frac{2 \cdot \sigma_0^2 \cdot \ln \dfrac{\sigma_1}{\sigma_0}}{1 - \left(\dfrac{\sigma_0}{\sigma_1}\right)^2} = \frac{2 \cdot 2^2 \cdot \ln \dfrac{2,5}{2}}{1 - \left(\dfrac{2}{2,5}\right)^2} = 4,959; \quad d = -\frac{\ln 0,1}{\ln \dfrac{2,5}{2}} = 10,319.$$

Тогда получаем $\tilde{y}_j^{\text{н}i} = y_j - 4,959 \cdot (j + 10,319) + 4,959 \cdot i$, $i < j$.

Результаты расчета контрольных границ $\tilde{y}_j^{\text{н}i}$ приведены в таблице.

j	Экспериментальные значения y_i								
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
	3,12	49,38	76,68	117,48	163,48	182,68	199,08	224,16	246,66
2	−1,79								
3	−53,80	−9,20							
4	−52,70	−8,08	26,63						
5	−31,50	13,13	47,84	87,51					
6	−32,10	12,49	47,12	86,86	111,71				
7	−35,57	9,05	47,77	83,44	108,23	128,07			
8	−40,24	4,38	39,09	78,76	103,56	123,40	143,24		
9	−42,50	2,09	44,63	79,34	119,02	143,81	163,65	193,40	
10	−42,30	2,33	37,05	76,72	101,51	121,35	141,18	170,94	195,73

Из таблицы видно, что все опытные значения лежат выше критической границы. Следовательно, смещение дисперсии вверх не обнаруживается.

5.4.2.4. ККНС для доли дефектных изделий

В качестве контролируемой величины используется $y_i = \sum_i x_i$, где x_i — число дефектных изделий в выборке объема n. Так как практический интерес представляет случай превышения некоторой нормативной доли дефектных изделий p_0, то будем рассматривать одностороннюю контрольную карту. Цель такой карты — своевременно зафиксировать превышение нормативной долей дефектных изделий p_0 ее допустимого уровня p_1.

Параметры такой контрольной карты вычисляются по формулам

$$b = \frac{n \ln \dfrac{1-p_0}{1-p_1}}{\ln \dfrac{p_1(1-p_0)}{p_0(1-p_0)}}; \quad d = -\frac{\ln \alpha}{n \ln \dfrac{1-p_0}{1-p_1}}.$$

Уравнения для контрольных границ имеют вид

$$\tilde{y}_j^{\text{н}i} = y_j - \frac{n \ln \dfrac{1-p_0}{1-p_1}}{\ln \dfrac{p_1(1-p_0)}{p_0(1-p_1)}} \left(+ \frac{\ln \alpha}{n \ln \dfrac{1-p_0}{1-p_1}} \right) + \frac{n \ln \dfrac{1-p_0}{1-p_1}}{\ln \dfrac{p_1(1-p_0)}{p_0(1-p_1)}} i, \quad i < j.$$

Если $y_i < \tilde{y}_j^{\text{н}i}$, то следует вывод о значимом росте доли дефектных изделий в производстве.

Задача 390. *В результате испытаний $m = 10$ партий изделий объема $n = 100$ каждая получены количества дефектных изделий в каждой выборке (данные приведены в таблице). Найти параметры ККНС для доли дефектных изделий при $\alpha = 0{,}1$, $p_0 = 0{,}02$ и $p_1 = 0{,}03$.*

Номер выборки, i	1	2	3	4	5	6	7	8	9	10
x_i	3	1	1	3	2	2	5	1	3	2
$\sum x_i$	3	4	5	8	10	12	17	18	21	23

Вычисляем параметры ККНС для доли дефектных изделий

$$b = \frac{100 \cdot \ln\left(\dfrac{1-0{,}02}{1-0{,}03}\right)}{\ln\left[\dfrac{0{,}03\cdot(1-0{,}02)}{0{,}02\cdot(1-0{,}03)}\right]} = 2{,}46716; \quad d = -\frac{\ln 0{,}1}{100\cdot\ln\left(\dfrac{1-0{,}02}{1-0{,}03}\right)} = 2{,}245.$$

Следовательно, $\tilde{y}_j^{\text{н}i} = y_j - 2{,}467\cdot(j+2{,}245) + 2{,}467\cdot i$.

Расчеты значений контрольных границ $\tilde{y}_j^{\text{н}i}$ сведем в таблицу.

j	Экспериментальные значения y_i								
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
	3	4	5	8	10	12	17	18	21
2	−4,00								
3	−5,47	−3,00							
4	−4,94	−2,47	−0,01						
5	−5,41	−2,94	−0,47	1,99					
6	−5,87	−3,41	−0,94	1,53	3,99				
7	−3,34	−0,87	1,59	4,06	6,53	8,99			
8	−4,80	−2,34	0,13	2,59	5,06	7,53	9,99		
9	−4,27	−1,81	0,66	3,13	5,59	8,06	10,53	12,99	
10	−4,74	−2,27	0,19	2,66	5,13	7,59	10,06	12,53	14,99

Из таблицы следует, что ни одно значение y_i не попадает в зону контроля, следовательно, нет оснований считать, что доля дефектных изделий в выборках отличается от нормативной.

5.4.2.5. ККНС для числа дефектных изделий, основанная на распределении Пуассона

Координаты точек ККНС те же, что и в предыдущем разделе $(i, \sum x_i)$. Параметры ККНС вычисляются по формулам

$$b = \frac{\mu - \mu_0}{\ln\dfrac{\mu_1}{\mu_0}}; \quad d = -\frac{\ln\alpha}{\mu_1 - \mu_0},$$

где μ_0 — нормативное значение числа дефектных изделий; μ_1 — сдвиг числа дефектных изделий, который должен быть обнаружен контролем; α — ошибка первого рода (рекомендуемые значения 0,01; 0,05; 0,1).

Уравнение для контрольной границы односторонней ККНС с контролем сдвига вверх числа дефектных изделий имеет вид

$$\tilde{y}_j^{\text{н}i} = y_j - b\,(j+d) + bi, \quad i < j.$$

Если $y_i < \tilde{y}_j^{\text{н}i}$ $(i < j)$, то увеличение количества дефектных изделий признается значимым.

Задача 391. *Результаты испытаний партии изделий приведены в таблице. Нормативное значение* $\mu_0 = 3$, *допустимое* $\mu_1 = 5$, $\alpha = 0{,}1$. *Найти параметры ККНС для числа дефектных изделий и провести анализ полученных результатов.*

Номер выборки, i	1	2	3	4	5	6	7	8	9	10
Число дефектов, x_i	3	4	2	2	3	4	9	3	2	2
$\sum x_i$	3	7	9	11	14	18	27	30	32	34

Вычисляем

$$b = \frac{5-3}{\ln\frac{5}{3}} = 3{,}915; \quad d = -\frac{\ln 0{,}1}{5-3} = 1{,}151.$$

Тогда $\tilde{y}_j^{\text{н}i} = y_j - 3{,}915 \cdot (j + 1{,}151) + 3{,}915 \cdot i$, $i < j$

Расчеты контрольных границ $\tilde{y}_j^{\text{н}i}$ сведем в таблицу.

j	Экспериментальные значения y_i								
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
	3	7	9	11	14	18	27	30	32
2	−1,42								
3	−3,33	0,58							
4	−5,25	−1,34	2,58						
5	−6,17	−2,25	1,66	5,58					
6	−6,08	−2,17	1,75	5,66	9,58				
7	−1,00	−2,92	6,83	10,75	**14,66**	**18,58**			
8	−1,91	2,00	5,92	9,83	13,75	17,66	21,58		
9	−3,82	0,09	4,00	7,92	11,83	15,75	19,66	23,58	
10	−5,74	−1,82	2,09	6,00	9,92	13,83	17,75	21,66	25,58

Из таблицы видим, что в седьмой выборке значимо выросло количество дефектных изделий ($y_5 = 15 < \bar{y}_7^{\text{н}} = 14{,}66$ и $y_6 = 18 < \tilde{y}_7^{\text{н}} = 18{,}58$). Применение ККНС позволило оперативно выявить этот сдвиг.

5.4.3. Относительная эффективность контрольных карт

У инженера производства может возникнуть закономерный вопрос: которую из контрольных карт следует применять? Вообще говоря, для того, чтобы дать более полный ответ на вопрос, следовало бы знать, чего же хочет инженер: упростить вычисления, обеспечить высокую эффективность контроля, какие изменения кон-

Таблица 241

Относительная эффективность различных контрольных карт [14]

Причина изменения	Карты Шухарта			ККНС
	$\bar{x}$-карта	R-карта	s-карта	
Отклонение с пересечением (грубое)	1	2	–	3
Сдвиг среднего	2	2	3	1
Сдвиг дисперсии	–	1	–	–
Медленная флуктуация (тренд)	2	–	–	1
Быстрая флуктуация (колебания)	–	1	2	–

тролируемого процесса хотел бы он зафиксировать? Ответы на эти вопросы может дать только консультация, проводимая специалистом с учетом всех особенностей и специфики решаемой задачи. Однако самые общие рекомендации можно дать — они представлены в табл. 241. В таблице приняты следующие оценки: 1 — наиболее эффективна; 2 — следующая по эффективности; 3 — наименее эффективна; 4 — (прочерк) неприменима.

5.4.4. Контроль без использования контрольных карт

Для администратора не являющегося статистиком, контрольные карты Шухарта, а тем более ККНС, являются трудно усвояемыми, так как требуют привлечения специальных терминов, вычислений по формулам, применения специальных таблиц. Поэтому были предприняты попытки создать простейшие методы статистического анализа, которые не требуют ни таблиц, ни вычислений.

Один из таких методов описан в [618]. Правила и логика метода состоят в следующем. Для непрерывного производства, т. е. когда можно говорить о производстве с определенной „настройкой“, оператор-контролер, работающий с контрольным устройством, снабжается следующей инструкцией. На контрольном устройстве отмечаются красным цветом зоны, соответствующие браку: для одностороннего контроля — одна зона, для двустороннего контроля — две. Далее на контрольном устройстве помечаются зеленым и желтым цветами зоны контроля в соответствии с диаграммой:

— для двустороннего контроля

	100%			
Красная зона	Желтая зона 25%	Зеленая зона 50%	Желтая зона 25%	Красная зона

— для одностороннего контроля

100%		
Зеленая зона 50%	Желтая зона 50%	Красная зона

Для контроля используется правило двух последовательных изделий. Если без изменения настройки производства последовательно получаем два „зеленых“ или одно „зеленое“ и одно „желтое“ изделие, то работа продолжается без изменений.

Однако всегда в начале смены и после поступления новой партии изделий работу следует начинать только после получения пяти последовательных „зеленых“ изделий. Если получается „желтый“ результат, следует начать подсчет „зеленых“ изделий заново, и так до тех пор, пока не будут получены пять последовательных „зеленых“ изделий.

Если после настройки в процессе работы получены подряд два „желтых“ результата на одной и той же стороне допуска, то требуется „регулировка в центр“ (это должен уметь делать оператор производства). Затем обязательна проверка по правилу — пять „зеленых“ подряд.

Появление двух „желтых“ по разные стороны от центра требует вмешательства руководителя производства (увеличилось рассеяние результата). После устранения причины этого увеличения требуется опять получить пять „зеленых“ изделий подряд.

Коротко повторим основные правила метода системы контроля стабильности производства:

— начинать с пяти „зеленых";

— продолжать работу всегда, если подряд получены два „зеленых" либо „зеленое" и „желтое" изделия;

— если получены подряд два „желтых" изделия, регулировать настройку производства;

— после регулировки снова требуется получить пять „зеленых" изделий подряд.

Естественно, такой метод контроля наиболее эффективен в поточном массовом производстве однотипных изделий, к производству которых применимо понятие „настройка".

5.5. Математико-статистические методы планирования эксперимента

Планирование эксперимента предполагает определение наиболее эффективной стратегии его проведения с целью получения статистического материала, обладающего заранее заданными свойствами.

С примером такого планирования мы уже встречались в разделе 5.3.3.2 (использование равноотстоящих значений независимой переменной при поиске нелинейной регрессии).

Планирование эксперимента применяется при решении таких задач, как оценка параметров распределения, проверка статистических гипотез при заданной мощности критерия, нахождение математической модели процесса с заданными статистическими свойствами, поиск оптимальных по заданным критериям условий протекания изучаемого процесса.

Многие из методов уже рассматривались нами в соответствующих разделах. В настоящем разделе мы уделим особое внимание математическим методам планирования эксперимента для изучения механизма наблюдаемого процесса и построения его статистической модели.

В последние годы математическая теория активного эксперимента, бурно развиваясь, оформилась в самостоятельное прикладное направление математической статистики [619, 620]. Автор далек от мысли изложить стройные основы теории планирования эксперимента. Скорее всего, он в состоянии предложить читателю введение в „прихожую" таких основ. Однако автор надеется дать основные понятия и изложить простейшие прикладные методы планирования эксперимента, вполне доступные специалистам с общим высшим образованием.

5.5.1. Планирование регрессионных экспериментов при изучении механизма явления (статистическое моделирование)

Задача планирования эксперимента может быть сформулирована следующим образом: требуется получить некоторое представление о поверхности отклика, описываемой моделью $y = f(x_1, x_2, \ldots, x_n)$, где y — зависимая переменная-отклик (например, долговечность изделия); x_i — независимые переменные — влияющие на отклик факторы, которые можно варьировать в ходе эксперимента (например, параметры режима нагружения изделия).

Неизвестная функция отклика чаще всего представляется полиномом k-й степени

$$y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{\substack{i,j=1 \\ i<j}}^{k} b_{ij} x_i x_j + \sum_{i=1}^{k} b_{ii} x_i^2 + \ldots,$$

где b_0, b_i, b_{ij}, b_{ii} — коэффициенты полинома.

Планирование эксперимента заключается в выборе на каждом этапе исследования оптимального в принятых критериях расположения экспериментальных точек в пространстве факторов. В качестве критериев оптимальности планов используются:

— минимизация числа опытов;

— простота вычислений коэффициентов функции отклика;

— независимость оценок коэффициентов функции отклика (ортогональность плана);

— однородность дисперсий отклика относительно центра плана (ротатабельность плана);

— минимизация объема эллипсоида рассеяния оценок коэффициентов модели (D-оптимальность плана).

Наибольшее применение нашли ортогональные планы в сочетании с критериями D-оптимальности.

Как правило, исследователь ищет поверхность отклика в какой-то определенной области изменения факторов. Наиболее широко применяется планирование на двух уровнях (экстремальный эксперимент), когда в эксперименте используются значения факторов, соответствующие верхней и нижней границам интервала его варьирования. Эти значения называются верхним и нижним уровнями и обозначаются $+1$ и -1 соответственно (или просто $+$ и $-$). Экспериментальные планы, в которых все факторы варьируются только на двух уровнях, называются планами 2^k, где k — число варьируемых факторов.

При построении плана эксперимента исследователь должен исходить из некоторого априорного представления о возможном виде функции отклика (линейность, монотонность и т. п.). Сначала область варьирования факторов определяется исходя из предположения о линейности поверхности отклика внутри этой области. Если линейная поверхность отклика описывает экспериментальный материал неадекватно, то проводятся дальнейшие эксперименты по уточнению ее вида с помощью полиномов более высокого порядка.

Более полное изложение математической теории планирования эксперимента содержится в [621–625].

5.5.1.1. Линейные ортогональные планы (планирование первого порядка)

5.5.1.1.1. Полный факторный эксперимент

Примером плана, позволяющего получить независимые оценки коэффициентов поверхности отклика, является полный факторный эксперимент (ПФЭ), реализующий все возможные неповторяющиеся комбинации уровней независимых факторов (ранее планы такого типа мы обсуждали применительно к задачам дисперсионного анализа).

В настоящем разделе ПФЭ рассматриваются применительно к условиям активного экстремального эксперимента.

Рассмотрим в качестве примера ПФЭ для трех факторов ($k = 3$). Поверхность отклика в этом случае имеет вид

$$y = b_0 + \sum_{i=1}^{3} b_i x_i + \sum_{\substack{i,j=1 \\ i<j}}^{k} b_{ij} x_i x_j + b_{123} x_1 x_2 x_3.$$

Пусть каждый фактор x_i варьируется от основного уровня x_{i0} на величину $\pm\Delta x_i$. Тогда с помощью преобразования $\tilde{x}_i = \dfrac{x_i - x_{i0}}{\Delta x_i}$ можно перейти к кодированным переменным $\tilde{x}_i$, принимающим на границах интервала варьирования x_i значения ± 1. План экстремального эксперимента принято записывать в виде матрицы, определяющей в кодированных переменных $\tilde{x}_i$ условия проведения эксперимента (в дальнейшем под x_i будем понимать кодированные переменные без специальных оговорок). Пример матрицы ПФЭ для трех факторов приведен в табл. 242. В литературе принято k-факторный ПФЭ с изменением факторов на двух уровнях называть планом типа 2^k.

В дальнейшем будем обозначать через $y_{j\nu}$ значение отклика, полученное в ν-м эксперименте при условиях (сочетании уровней факторов x_i), соответствующих j-й точке плана, и через x_{ij} — значение фактора x_i в j-й точке плана.

Таблица 242

Матрица полного факторного эксперимента (ПФЭ) 2^3

Номер точки плана, j	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	Отклик y_j
1	+	−	−	−	+	+	+	−	y_1
2	+	+	−	−	−	−	+	+	y_2
3	+	−	+	−	−	+	−	+	y_3
4	+	+	+	−	+	−	−	−	y_4
5	+	−	−	+	+	−	−	+	y_5
6	+	+	−	+	−	+	−	−	y_6
7	+	−	+	+	−	−	+	−	y_7
8	+	+	+	+	+	+	+	+	y_8

Для ПФЭ имеют место соотношения ($n = 2^k$ — число точек плана):

$$\sum_{j=1}^{n} x_{ij} = 0 \ (i = 1, 2, \ldots, 2^{k-1}); \qquad \sum_{j=1}^{n} x_{ij}^2 = n;$$

$$\sum_{j=1}^{n} x_{ij}x_{\varepsilon j} = 0 \quad (i \neq \varepsilon; \ i, \varepsilon = 0, 1, \ldots, 2^{k-1}),$$

что соответствует свойству ортогональности столбцов матрицы плана.

Любой план 2^k может быть построен по следующему простому правилу: в столбце, соответствующем фактору x_i, знаки + и − чередуются через 2^{i-1}. План 2^k позволяет оценить 2^k коэффициентов регрессии b_i. Однако использовать ПФЭ для оценки коэффициентов при членах с кратностью нельзя, так как оценки для b_0 и b_{ii} смешиваются (например, столбцы x_0 и x_ix_i неразличимы).

Основным преимуществом ПФЭ является ортогональность матрицы плана, что позволяет существенно упростить вычисление коэффициентов уравнения отклика. Для любого числа факторов k выборочные оценки b_i вычисляются по формулам [621]

$$b_i = \frac{1}{n}\sum_{i=1}^{n} b_{ij}\bar{y}_j, \quad \text{где} \quad \bar{y}_j = \frac{1}{m}\sum_{\nu=1}^{m} y_{j\nu};$$

m — число параллельных опытов в j-й точке плана; n — общее число точек плана.

Дисперсия, характеризующая разброс значений $y_{j\nu}$ при постоянных условиях эксперимента (т. е. в одной точке плана), находится по формуле $S_j^2 = \dfrac{1}{m-1}\sum\limits_{\nu=1}^{m} (y_{j\nu} - \bar{y}_j)^2$. Общая дисперсия, характеризующая разброс отклика безотносительно к условиям эксперимента, равна $S_y^2 = \dfrac{1}{n}\sum\limits_{j=1}^{n} S_j^2 = \dfrac{1}{n(m-1)}\sum\limits_{j=1}^{n}\sum\limits_{\nu=1}^{m} (y_{j\nu} - \bar{y}_j)^2$.

Если количество параллельных опытов в точках плана различно, то $S_y^2 = \dfrac{\sum\limits_{j=1}^{n}(m_j - 1)\,S_j^2}{\sum\limits_{j=1}^{n}(m_j - 1)}$, где m_j — число опытов в j-й точке плана.

Предварительно однородность дисперсий S_j^2 должна быть проверена одним из методов, изложенных в разделе 4.1.1.4.

Дисперсия коэффициента регрессии b_i определяется формулой

$$S_{b_i}^2 = \frac{1}{n\,(m-1)} S_y^2.$$

Коэффициент b_i уравнения отклика с достоверностью α признается значимым, если $|b_i| > t_{\frac{1+\alpha}{2}}(f)\,S_{b_i}$, где $t_{\frac{1+\alpha}{2}}$ — квантиль распределения Стьюдента при $f = n(m-1)$ степенях свободы.

Для проверки адекватности математической модели отклика используется дисперсия $S^2 = \dfrac{m}{n-d}\sum\limits_{j=1}^{n}(\bar{y}_j - \hat{y}_j)^2$, где d — число коэффициентов аппроксимирующего полинома; $\hat{y}_j$ — значение отклика, предсказываемое регрессионной моделью. Адекватность модели устанавливается сравнением дисперсий S^2 и S_y^2 с помощью критерия Фишера $F = \dfrac{S^2}{S_y^2}$ при $f_1 = n - d$ и $f_2 = n\,(m-1)$ степенях свободы.

Если все коэффициенты линейной регрессии (в том числе и все коэффициенты при взаимодействиях) являются значимыми, то $d = n$ и не остается степеней свободы для проверки гипотезы адекватности. В этом случае рекомендуется поставить эксперименты в центре плана (т. е. при значении фактора $x_i = 0$). Тогда, если $|b_0 - \bar{y}_0| < S_y$, где $\bar{y}_0$ — среднее значение отклика в центре эксперимента, линейная модель признается адекватной.

Задача 392. *Построить матрицу ПФЭ* 2^5.

При построении матрицы плана эксперимента будем исходить из правила чередования для фактора x_i знаков $+$ и $-$ через каждые 2^{i-1} знаков. Например, для фактора x_4 знаки должны чередоваться через $2^{4-1} = 8$ знаков. Следуя этому правилу, получаем матрицу плана эксперимента ПФЭ 2^5:

j	x_0	x_1	x_2	x_3	x_4	x_5	j	x_0	x_1	x_2	x_3	x_4	x_5
1	+	+	+	+	+	+	17	+	+	+	+	+	−
2	+	−	+	+	+	+	18	+	−	+	+	+	−
3	+	+	−	+	+	+	19	+	+	−	+	+	−
4	+	−	−	+	+	+	20	+	−	−	+	+	−
5	+	+	+	−	+	+	21	+	+	+	−	+	−
6	+	−	+	−	+	+	22	+	−	+	−	+	−
7	+	+	−	−	+	+	23	+	+	−	−	+	−
8	+	−	−	−	+	+	24	+	−	−	−	+	−
9	+	+	+	+	−	+	25	+	+	+	+	−	−
10	+	−	+	+	−	+	26	+	−	+	+	−	−
11	+	+	−	+	−	+	27	+	+	−	+	−	−
12	+	−	−	+	−	+	28	+	−	−	+	−	−
13	+	+	+	−	−	+	29	+	+	+	−	−	−
14	+	−	+	−	−	+	30	+	−	+	−	−	−
15	+	+	−	−	−	+	31	+	+	−	−	−	−
16	+	−	−	−	−	+	32	+	−	−	−	−	−

Задача 393. *Для ПФЭ 2^3, приведенного в таблице, найти уравнение отклика и провести его статистический анализ.*

Уровни факторов и интервалы варьирования

Уровни факторов	Обозначение	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$
Основной	0	7	2	4
Интервал варьирования	Δx_i	4	2	3
Верхний	+1	11	4	7
Нижний	−1	3	0	1

Матрица ПФЭ 2^3 и результаты эксперимента

j	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y_j	$\bar{y}_j$	$\bar{S}_j^2$	$\tilde{y}_j$
1	+	+	+	+	+	+	+	+	34, 38, 42	38	16	37,5
2	+	−	+	+	−	−	+	−	15, 20, 25	20	25	18,5
3	+	+	−	+	−	+	−	−	19, 21, 23	21	4	21,5
4	+	−	−	+	+	−	−	+	30, 33, 36	33	9	34,5
5	+	+	+	−	+	−	−	−	−33, −38, −43	−38	25	−38,5
6	+	−	+	−	−	+	−	+	28, 31, 34	31	9	29,5
7	+	+	−	−	−	−	+	+	−19, −23, −27	−23	16	−22,5
8	+	−	−	−	+	+	+	−	10, 12, 14	12	4	13,5

Вычисляем коэффициенты регрессии (уравнения отклика)

$$b_0 = \frac{1}{8} \cdot \sum_{j=1}^{8} x_{0j} \cdot \bar{y}_j = \frac{94}{7} = 11{,}75;$$

$$b_1 = \frac{1}{8} \cdot \sum_{j=1}^{8} x_{1j} \cdot \bar{y}_j = \frac{38 - 20 + 21 - 33 - 38 - 31 - 23 - 12}{8} = -12{,}25;$$

$$b_2 = \frac{38 + 20 - 21 - 33 - 38 + 31 + 23 - 12}{8} = 1{,}0;$$

$$b_3 = \frac{38 + 20 + 21 + 33 + 38 - 31 + 23 - 12}{8} = 16{,}25;$$

$$b_{12} = \frac{1}{8} \cdot \sum_{j=1}^{8} x_{1j} \cdot x_{2j} \cdot \bar{y}_j = \frac{38 - 20 - 21 + 33 - 38 - 31 + 23 + 12}{8} = -0{,}5;$$

$$b_{13} = \frac{38 - 20 + 21 - 33 + 38 + 31 + 23 + 12}{8} = 13{,}75;$$

$$b_{23} = \frac{38 + 20 - 21 - 33 + 38 - 31 - 23 + 12}{8} = 0;$$

$$b_{123} = \frac{1}{8} \cdot \sum_{j=1}^{8} x_{1j} \cdot x_j \cdot x_{3j} \cdot \bar{y}_j = \frac{38 - 20 - 21 + 33 + 38 + 31 - 23 - 12}{8} = 8{,}0.$$

Далее вычисляем $\bar{S}_j^2$ (значения приведены в таблице) и

$$S_y^2 = \frac{1}{8} \cdot \sum_{j=1}^{8} \bar{S}_j^2 = \frac{16 + 25 + 4 + 9 + 25 + 16 + 4}{8} = 13{,}5.$$

Находим дисперсию коэффициентов регрессии:

$$S_{b_i}^2 = \frac{1}{n \cdot (m-1)} \cdot S_y^2 = \frac{13{,}5}{8 \cdot (3-1)} = 0{,}844 \quad (S_{b_i} = 0{,}918)\,.$$

Выбираем $\alpha = 0{,}95$ и находим из таблиц $t_{\frac{1+\alpha}{2}} = t_{0,975}$ при $f = 8 \cdot (3-1) = 16$ степенях свободы: $t_{0,975}(16) = 2{,}12$. Вычисляем статистики Стьюдента для коэффициентов модели:

$$t_0 = \frac{|b_0|}{S_{b_i}} = \frac{11{,}75}{0{,}918} = 12{,}8; \quad t_1 = \frac{|b_1|}{S_{b_i}} = \frac{12{,}25}{0{,}918} = 13{,}3; \quad t_2 = \frac{|b_2|}{S_{b_i}} = \frac{1{,}0}{0{,}918} = 1{,}09;$$

$$t_3 = \frac{|b_3|}{S_{b_i}} = \frac{16{,}25}{0{,}918} = 17{,}7; \quad t_{12} = \frac{|b_{12}|}{S_{b_i}} = \frac{0{,}5}{0{,}918} = 0{,}54; \quad t_{13} = \frac{|b_{13}|}{S_{b_i}} = \frac{13{,}75}{0{,}918} = 14{,}98;$$

$$t_3 = \frac{|b_{23}|}{S_i} = \frac{0}{0{,}918} = 0; \quad t_{123} = \frac{|b_{123}|}{S_{b_i}} = \frac{8{,}0}{0{,}918} = 8{,}71.$$

Видим, что $t_2 = 1{,}09$; $t_{12} = 0{,}54$; $t_{23} = 0 < t_{0,975}(16) = 2{,}12$. Следовательно, коэффициенты b_2, b_{12} и b_{23} незначимо отличаются от нуля. Тогда уравнение регрессии принимает вид

$$y = 11{,}75 - 12{,}25 \cdot x_1 + 16{,}25 \cdot x_3 + 13{,}75 \cdot x_1x_3 + 8 \cdot x_1x_2x_3.$$

Вычислим теперь $\sum_{j=1}^{8} (\bar{y}_j - \hat{y}_j)^2 = 10$. В нашем случае число значимых коэффициентов модели равно $d = 5$ и $S^2 = \frac{m}{n-d} \cdot \sum_{j=1}^{8} (\bar{y}_j - \hat{y}_j)^2 = \frac{3 \cdot 10}{8-5} = 10$.

Теперь находим $F = \frac{S^2}{S_y^2} = \frac{10}{13{,}5} = 0{,}74$, что, очевидно, меньше критического значения (которое всегда больше единицы). Запишем искомую модель в натуральных переменных:

$$y = 11{,}75 - 12{,}25 \cdot \left(\frac{x_1-7}{4}\right) + 16{,}25 \cdot \left(\frac{x_3-4}{3}\right) + 13{,}75 \cdot \left(\frac{x_1-7}{4}\right) \cdot \left(\frac{x_3-4}{3}\right) +$$
$$+ 8 \cdot \left(\frac{x_1-7}{4}\right) \cdot \left(\frac{x_2-2}{2}\right) \cdot \left(\frac{x_3-4}{3}\right) = 11{,}75 - 12{,}25 \cdot (0{,}25x_1 - 1{,}75) + 16{,}25 \cdot (0{,}33x_3 - 1{,}33) +$$
$$+ 13{,}75 \cdot (0{,}25x_1 - 1{,}75) \cdot (0{,}33x_3 - 1{,}33) + 8 \cdot (0{,}25x_1 - 1{,}75) \cdot (0{,}5x_2 - 1) \cdot (0{,}33x_3 - 1{,}33) =$$
$$= 24{,}958 - 4{,}974 \cdot x_1 + 9{,}31 \cdot x_2 + 2{,}042 \cdot x_3 - 1{,}33 \cdot x_1 \cdot x_2 +$$
$$+ 0{,}473 \cdot x_1 \cdot x_3 - 2{,}31 \cdot x_2 \cdot x_3 + 0{,}33 \cdot x_1 \cdot x_2 \cdot x_3.$$

5.5.1.1.2. Дробный факторный эксперимент

Если некоторыми взаимодействиями можно пренебречь, то регрессионная модель, связывающая отклик с основными факторами, может быть получена при меньшем, чем в ПФЭ, количестве экспериментов с помощью дробного факторного эксперимента (ДФЭ).

Например, в случае ПФЭ 2^3, если взаимодействиями x_1x_2, x_1x_3, x_2x_3 и $x_1x_2x_3$ можно пренебречь, то можно либо использовать 4 оставшихся столбца матрицы плана для определения коэффициентов регрессионной модели при четырех новых факторах, либо найти модели для трех факторов с помощью четырех, а не восьми экспериментов. Часть матрицы ПФЭ 2^k, в которой ν линейных эффектов (факторов) приравнены к эффектам взаимодействия, называется дробной репликой вида $2^{k-\nu}$.

Соотношения, определяющие правила построения дробных реплик ПФЭ, и указывающие, какие факторы приравнены к взаимодействиям, называются генерирующими. Например, дробные реплики ПФЭ 2^{3-1} могут быть получены с помощью генерирующих соотношений $x_3 = x_1x_2$ и $x_3 = -x_1x_2$. Матрицы планов, соответствующих этим дробным репликам, приведены в табл. 243.

Очевидно, можно получить еще 4 матрицы дробных реплик 2^{3-1} с помощью генерирующих соотношений для факторов x_1 и x_2.

Таблица 243

Дробные реплики 2^{3-1}

$x_3 = x_1x_2$					$x_3 = -x_1x_2$				
j	x_1	x_2	x_3	$x_1x_2x_3$	j	x_1	x_2	x_3	$x_1x_2x_3$
1	+	+	+	+	1	+	+	−	−
2	−	−	+	+	2	−	−	−	−
3	+	−	−	+	3	+	−	+	−
4	−	+	−	+	4	−	+	+	−

Для различения смешанных эффектов по матрице дробной реплики используется понятие определяющего контраста, характеризующего комбинацию тех факторов, столбец произведения которых состоит только из плюсов или только из минусов.

Определяющий контраст J может быть равен $+1$ или -1. Например, для дробных реплик (табл. 243)

$$J = x_1x_2x_3 = +1 \quad \text{или} \quad J = x_1x_2x_3 = -1.$$

Определяющий контраст позволяет установить систему смешивания основных факторов с эффектами взаимодействия. Например, если $J = x_1x_2x_3 = +1$, то $x_1 = x_1^2x_2x_3 = x_2x_3$ (так как всегда $x_i^2 = +1$), т. е. оценка b_1 смешана с оценкой b_{23}.

Если к эффектам взаимодействия приравнивается не один, а несколько основных факторов, причем каждому из них соответствует свой частный определяющий контраст, для полного описания разрешающей способности дробной реплики ПФЭ используется обобщающий определяющий контраст. Он включает в себя частные определяющие контрасты и их произведения.

Например, при исследовании 5 факторов можно поставить не $2^5 = 32$ опыта, а только 8, если реализовать дробную реплику 2^{5-2}, т. е. приравнять два фактора к эффектам взаимодействия. Предположим, что выбраны варианты смешивания $x_4 = x_1x_3$ и $x_5 = x_1x_2x_3$ с определяющими контрастами $J = x_1x_3x_4$ и $J = x_1x_2x_3x_5$ соответственно. Тогда обобщающий определяющий контраст может быть записан следующим образом

$$J = x_1x_3x_4 = x_1x_2x_3x_5 = x_2x_4x_5.$$

Теперь для того, чтобы выявить, с чем смешана та или иная оценка коэффициента модели, необходимо умножить комбинацию факторов, соответствующих ей, на обобщающий определяющий контраст. Например, определим систему смешивания для оценки коэффициента b_{12} эффекта взаимодействия x_1x_2. Имеем

$$x_1x_2 = x_2x_3x_4 = x_3x_5 = x_1x_4x_5,$$

т. е. оценка b_{12} будет смешанной с оценкой коэффициентов b_{234}, b_{35}, b_{145}. В общем случае получается сложная система смешивания эффектов взаимодействия.

Наиболее эффективными дробными репликами от ПФЭ являются реплики, у которых линейные эффекты смешаны с взаимодействиями наивысшего порядка.

Различают регулярные и нерегулярные дробные реплики. Регулярные реплики образуются из ПФЭ делением на число частей, кратное двум. Например, 2^{5-1} — полуреплика, 2^{5-2} — четверть реплики, 2^{7-4} — 1/16 реплики от ПФЭ 2^7. Реплики типа 3/4, 5/8 и т. д. называются нерегулярными. Дробные реплики позволяют существенно сократить число факторов и экспериментов для моделирования процесса. Особенно эффективно их применение при планировании экспериментов для отыскания оптимума отклика.

5.5.1.2. Нелинейные планы второго порядка

Если линейная регрессионная модель оказывается неадекватной, то в большинстве практических случаев удовлетворительная аппроксимация поверхности отклика достигается при использовании полинома второй степени. Математическая модель поверхности отклика в этом случае имеет вид

$$y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i<j}^{k} b_{ij} x_i x_j + \sum_{i=1}^{k} b_{ii} x_i^2.$$

План эксперимента должен обеспечивать получение раздельных (несмешанных) оценок коэффициентов регрессии b_i. Для этого необходимо, чтобы между числом опытов плана n и числом коэффициентов модели k выполнялось соотношение $n > \frac{(k+2)(k+1)}{2}$, и чтобы каждый фактор варьировался не менее, чем на трех уровнях.

5.5.1.2.1. Симметричные планы второго порядка

Наиболее широкое распространение нашли симметричные планы второго порядка [621]. Под симметричным планом понимается план, удовлетворяющий соотношениям

$$\sum_{j=1}^{n} x_{ij} = 0; \quad \sum_{j=1}^{n} x_{\varepsilon i} x_{\gamma i} x_{\nu i} = 0 \ \ (\varepsilon, \gamma, \nu = 1, 2, \ldots, k; \ \varepsilon \neq \gamma \neq \nu \neq \varepsilon);$$

$$\sum_{i=1}^{n} x_{\varepsilon i} x_{\gamma i} = 0 \ \ (\varepsilon \neq \gamma); \quad \sum_{i=1}^{n} x_{\varepsilon j}^2 x_{\gamma i} = 0 \ \ (\gamma \neq \nu); \quad \sum_{i=1}^{n} x_{ij}^2 = n\lambda_2;$$

$$\sum_{i=1}^{n} x_{\varepsilon j}^2 x_{\gamma j}^2 = n\lambda_3 \ \ (\varepsilon \neq \gamma); \quad \sum_{i=1}^{n} x_{ij}^4 = n\lambda_4 \quad (\varepsilon, \gamma, \nu = 1, 2, \ldots, k),$$

где n — общее число опытов, выполняемое по плану; j — порядковый номер опыта.

Если в каждой точке плана проводится m_j параллельных опытов, то должно иметь место

$$n\lambda_2 = \sum_{j=1}^{\tilde{n}} m_j x_{ij}^2; \quad n\lambda_3 = \sum_{j=1}^{\tilde{n}} m_j x_{\varepsilon j}^2 x_{\gamma j}^2 \ \ (\varepsilon \neq \gamma);$$

$$n\lambda_4 = \sum_{j=1}^{\tilde{n}} m_j x_{ij}^4 \quad (\varepsilon, \gamma = 1, 2, \ldots, k),$$

где $\tilde{n}$ и $n = \sum_{i=1}^{\tilde{n}} m_j$ — соответственно число точек плана и общее число опытов.

Оценки коэффициентов регрессии для симметричного плана второго порядка вычисляются по формулам

$$b_0 = \frac{a}{n} \sum_{j=1}^{\tilde{n}} m_j \bar{y}_j - \frac{b}{n} \sum_{i=1}^{k} \sum_{j=1}^{\tilde{n}} m_j x_{ij}^2 \bar{y}_j;$$

$$b_{ij} = -\frac{b}{n} \sum_{j=1}^{\tilde{n}} m_j \bar{y}_j + \frac{c}{n} \sum_{j=1}^{\tilde{n}} m_j x_{ij}^2 \bar{y}_j - \frac{d}{n} \sum_{j=1}^{\tilde{n}} m_j x_{ij}^2 \bar{y}_j;$$

$$b_i = \frac{1}{\lambda_2 n} \sum_{j=1}^{\tilde{n}} m_j x_{ij} \bar{y}_j \ \ (i \neq 0); \quad b_{i\varepsilon} = \frac{1}{\lambda_3 n} \sum_{j=1}^{\tilde{n}} m_j x_{ij} x_{\varepsilon j} \bar{y}_j \ \ (i \neq \varepsilon),$$

где $\bar{y}_j = \dfrac{1}{m_j}\sum\limits_{\nu=1}^{m_j} y_{j\nu}$; $a = \dfrac{k\lambda_2^2}{\lambda_4 - \lambda_3 + k\lambda_3 - k\lambda_2^2} + 1$; $c = \dfrac{1}{\lambda_4 - \lambda_3}$;

$$b = \frac{\lambda_2}{\lambda_4 - \lambda_3 + k\lambda_3 - k\lambda_2^2}; \quad d = \frac{\lambda_3 - \lambda_2^2}{(\lambda_4 - \lambda_3)\left(\lambda_4 - \lambda_3 k\lambda_3 - k\lambda_2^2\right)}.$$

Контроль правильности вычислений оценок коэффициентов регрессии может быть выполнен с помощью равенства $\sum\limits_{j=1}^{n}(y_j - \hat{y}_j) = 0$, где $\hat{y}_j$ — предсказываемое моделью значение отклика.

Выборочные дисперсии и ковариации оценок коэффициентов модели определяются формулами

$$S_{b_0}^2 = \frac{a}{n}S_y^2; \quad S_{b_{ii}}^2 = \frac{c-d}{n}S_y^2; \quad S_{b_i}^2 = \frac{1}{\lambda_2 n}S_y^2; \quad S_{b_{ij}}^2 = \frac{1}{\lambda_3 n}S_y^2 \quad (i \neq j);$$

$$\operatorname{cov}(b_0, b_{ii}) = -\frac{b}{n}S_y^2; \quad \operatorname{cov}(b_{ii}, b_{jj}) = -\frac{d}{n}S_y^2 \quad (i \neq j),$$

где S_y^2 — выборочная дисперсия отклика y, связанная с ошибкой эксперимента.

Значение S_y^2 находится по формуле

$$S_y^2 = \frac{\sum\limits_{j=1}^{\tilde{n}}(m_j - 1)\,S_j^2}{\sum\limits_{j=1}^{\tilde{n}}(m_j - 1)} = \frac{\sum\limits_{j=1}^{\tilde{n}}\sum\limits_{\nu=1}^{m_j}(y_{j\nu} - \bar{y}_j)^2}{\sum\limits_{j=1}^{\tilde{n}}(m_j - 1)}, \quad \text{где} \quad S_j^2 = \frac{1}{m_j - 1}\sum_{\nu=1}^{m_j}(y_{j\nu} - \bar{y}_j)^2,$$

что справедливо в случае однородности дисперсий S_j^2 по точкам плана (эта однородность должна быть предварительно проверена).

Дисперсия $S_{\hat{y}}^2$ предсказываемого моделью значения отклика $\hat{y}$ определяется формулой

$$S_{\hat{y}}^2 = \frac{S_y^2}{n}\left[(c-d)\sum_{i=1}^{k} x_i^4 + \left(\frac{1}{\lambda_3} - 2d\right)\sum_{i<j}^{k} x_i^2 x_j^2 + \left(\frac{1}{\lambda_2} - 2b\right)\sum_{i=1}^{k} x_i^2 + a\right].$$

Неадекватность модели второго порядка характеризуется дисперсией $S^2 = \dfrac{1}{n-1}\sum\limits_{j=1}^{\tilde{n}} m_j\left(\bar{y_j}^{-}\,\hat{y}_j\right)^2$, где l — число коэффициентов модели (в случае квадратичной модели $l = \dfrac{(k+2)(k+1)}{2}$).

Модель с вероятностью α признается адекватной, если $\dfrac{S^2}{S_y^2} < F_\alpha$, где F_α — α-квантиль распределения Фишера с $f_1 = n - l$ и $f_2 = \sum\limits_{j=1}^{\tilde{n}}(m_j - 1)$ степенями свободы.

Значимость коэффициентов модели проверяется с помощью статистики $t_i = \dfrac{|b_i|}{S_{b_i}}$, имеющей при $b_i = 0$ распределение Стьюдента с $f = \sum\limits_{j=1}^{\tilde{n}}(m_j - 1)$ степенями свободы.

5.5.1.2.2. Ортогональные симметричные планы

Наиболее простыми симметричными планами второго порядка являются ортогональные планы, предложенные Боксом и Уилсоном [626]. Ортогональный план позволяет найти независимые оценки коэффициентов поверхности отклика. Построить ортогональный план для получения модели

$$\hat{y} = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i<j}^{k} b_{ij} x_i x_j + \sum_{i=1}^{k} b_{ii} x_i^2$$

нельзя, так как столбцы матрицы плана, соответствующие x_i^2 и фиктивной переменной x_0, будут неразличимы (оба будут состоять из $+1$) и, следовательно, неразличимы будут также оценки b_0 и b_{ii}. Для устранения этого смешивания используется преобразование модели к виду [621]

$$\eta = b_0' + \sum_{i=1}^{k} b_i x_i + \sum_{i<j}^{k} b_{ij} x_i x_j + \sum_{i=1}^{k} b_{ii} \left(x_i'\right)^2;$$

$$\text{где} \quad b_0' = b_0 + \lambda_2 \sum_{i=1}^{n} b_{ii}; \quad \left(x_i'\right)^2 = x_i^2 - \lambda_2;$$

$$\lambda_2 = \frac{1}{n} \sum_{\nu=1}^{n} x_{ij}^2 \quad \left(i = 1, 2, \ldots, k; \ \sum_{j=1}^{n} \left(x_{ij}'\right)^2 = 0\right).$$

Условие ортогональности плана имеет вид $\lambda_3 = \lambda_2^2$. Оценки коэффициентов регрессии на основе ортогонального плана получаются независимыми и находятся по формулам:

$$b_0' = \frac{1}{n} \sum_{j=1}^{n} y_j = \frac{1}{n} \sum_{j=1}^{\tilde{n}} m_j \bar{y}_j;$$

$$b_{ii} = \frac{\sum_{j=1}^{n} \left(x_{ij}'\right)^2 y_j}{\sum_{j=1}^{n} \left(x_{ij}'\right)^4} = \frac{1}{n\left(\lambda_4 - \lambda_3\right)} \sum_{j=1}^{n} x_{ij}^2 = \frac{1}{n\left(\lambda_4 - \lambda_3\right)} \sum_{j=1}^{\tilde{n}} m_j x_{ij}^2 \bar{y}_j;$$

$$b_i = \frac{1}{n\lambda_2} \sum_{j=1}^{n} x_{ij} y_j = \frac{\sum_{j=1}^{\tilde{n}} m_j x_{ij} \bar{y}_j}{\sum_{j=1}^{n} x_{ij}^2} \quad (i \neq 0);$$

$$b_{i\varepsilon} = \frac{1}{n\lambda_3} \sum_{j=1}^{n} x_{ij} x_{\varepsilon j} y_j = \frac{\sum_{j=1}^{\tilde{n}} m_j x_{ij} x_{\varepsilon j} \bar{y}_j}{\sum_{j=1}^{n} x_{ij}^2 x_{\varepsilon j}^2} \quad (i \neq \varepsilon).$$

Дисперсии оценок коэффициентов регрессии вычисляются по формулам

$$S_{b_0}^2 = \frac{1}{n} S_y^2; \quad S_{b_{ii}}^2 = \frac{1}{n(\lambda_4 - \lambda_3)} S_y^2; \quad S_{b_i}^2 = \frac{1}{n\lambda_2} S_y^2; \quad S_{b_{ij}}^2 = \frac{1}{n\lambda_3} S_y^2 \quad (i \neq j).$$

Оценка коэффициента b_0 находится по формуле

$$b_0 = b_0' - \lambda_2 \sum_{i=1}^{k} b_{ii};$$

ее выборочная дисперсия равна

$$S_{b_0}^2 = S_{b_0'}^2 + \lambda_2^2 \sum_{i=1}^{k} S_{b_{ii}}^2 = \frac{S_y^2 \left(\lambda_4 - \lambda_3 + k\lambda_2^2\right)}{n\left(\lambda_4 - \lambda_3\right)}.$$

Дисперсия предсказываемого значения отклика определяется по формуле

$$S_{\hat{y}}^2 = \frac{S_y^2}{n}\left[c\sum_{i=1}^{k} x_i^4 + \frac{1}{\lambda_3}\sum_{i<j1}^{k} x_i^2 x_j^2 + \left(\frac{1}{\lambda_2} - 2b\right)\sum_{i=1}^{k} x_i^2 + a\right].$$

Проверка значимости коэффициентов регрессии и адекватности поверхности отклика проводится по аналогии с обычным симметричным планом.

Большинство известных ортогональных планов второго порядка строится по композиционному принципу, путем достройки ортогонального плана первого порядка. Сначала реализуются опыты, соответствующие полуреплике 2^{3-1}, и опыты в центре плана для проверки линейности модели. Если модель неадекватна, ставятся опыты во второй полуреплике и в 6 звездных точках. Условие ортогональности $\lambda_3 = \lambda_2^2$ обеспечивается специальным выбором числа опытов в центре эксперимента n_0 и величиной звездного плеча α, определяющего координаты звездных точек плана. Композиционный ортогональный план содержит $n = 2^{k-p} + 2^k + n_0$ опытов, где 2^{k-p} — число опытов в звездных точках (дробная реплика ПФЭ, в котором p линейных эффектов приравнены к эффектам взаимодействия). Для композиционного плана имеют место соотношения

$$n\lambda_2 = \sum_{j=1}^{n} x_{ij}^2 = 2^{k-p} + 2\alpha^2; \quad n\lambda_3 = \sum_{j=1}^{n} x_{ij}^2 x_{\varepsilon j}^2 = 2^{k-p}; \quad n\lambda_4 = 2^{k-p} + 2\alpha^4.$$

Значения n_0 и α, обеспечивающие ортогональность плана, удовлетворяют соотношению

$$\left(2^{k-p} + 2k + n_0\right)2^{k-p} = \left(2^{k-p} + 2\alpha^2\right)^2.$$

Параметры и численные значения вспомогательных коэффициентов для ортогональных композиционных планов второго порядка при $n_0 = 1$ приведены в табл. 244. Вычисление коэффициентов модели и статистический анализ поверхности отклика проводятся по формулам для симметричного плана второго порядка.

Таблица 244

Параметры ортогональных композиционных планов

k	p	α^2	Распределение опытов			
			n	n_0	Вершина куба	Звездные точки
2	0	1,000	9	1	4	4
3	0	1,477	15	1	8	6
4	0	2,000	25	1	16	8
5	1	2,392	27	1	16	10

k	$\tilde{n}\lambda_2$	$\tilde{n}\lambda_3$	$\tilde{n}\lambda_4$	a	b	c	d
2	6,000	4,000	6,000	5,000	3,000	4,500	0
3	10,954	8,000	12,364	6,499	2,510	3,437	0
4	20,000	16,000	24,000	9,000	2,500	3,125	0
5	20,785	16,000	27,446	7,989	1,816	2,359	0

Задача 394. *Для трех переменных был реализован ортогональный план второго порядка, результаты которого приведены в таблице. Найти статистическую модель, описывающую результаты опытов.*

Уровни факторов

Уровни	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$
+1	15,1	36	1,4
0	14,7	32	1,2
−1	14,3	28	1,0

Из табл. 244 для $n_0 = 1$ (один опыт в центре плана) и $k = 3$ имеем величину звездного плеча $\alpha = \sqrt{1{,}477} = 1{,}215$. Общее число точек плана $\tilde{n} = 15$, число опытов в каждой точке плана было $m_j = 3$ (т. е. $n = \tilde{n} \cdot m_j = 45$). План состоит из 8 точек ПФЭ 2^3, 6 звездных точек и одной точки в центре плана (см. табл. 244).

Далее имеем

$$\lambda_2 = \frac{10{,}954}{15} = 0{,}73; \quad \lambda_3 = \frac{8}{15} = 0{,}53; \quad \lambda_4 = \frac{12{,}364}{15} = 0{,}824.$$

Матрица плана эксперимента имеет вид:

j	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	$x_{12} - \lambda^2$	$x_{22} - \lambda^2$	$\bar{y}_j$
1	−1	−1	−1	+1	+1	+1	0,27	0,27	0,27	16
2	+1	−1	−1	−1	−1	+1	0,27	0,27	0,27	5
3	−1	+1	−1	−1	+1	−1	0,27	0,27	0,27	9
4	+1	+1	−1	+1	−1	−1	0,27	0,27	0,27	5
5	−1	−1	+1	+1	−1	−1	0,27	0,27	0,27	14
6	+1	−1	+1	−1	+1	−1	0,27	0,27	0,27	10
7	−1	+1	+1	−1	−1	+1	0,27	0,27	0,27	17
8	+1	+1	+1	+1	+1	+1	0,27	0,27	0,27	9
9	−1,215	0	0	0	0	0	0,75	−0,73	−0,73	14
10	+1,215	0	0	0	0	0	0,75	−0,73	−0,73	8
11	0	−1,215	0	0	0	0	−0,73	0,75	−0,73	11
12	0	+1,215	0	0	0	0	−0,73	0,75	−0,73	8
13	0	0	−1,215	0	0	0	−0,73	−0,73	0,75	16
14	0	0	+1,215	0	0	0	−0,73	−0,73	0,75	16
15	0	0	0	0	0	0	−0,73	−0,73	−0,73	17

Вычисляем коэффициенты модели:

$$b_0' = \frac{1}{n} \cdot \sum_{j=1}^{n} \bar{y}_j = \frac{3}{45} \cdot \sum_{j=1}^{15} \bar{y}_j = 11{,}67;$$

$$b_{11} = \frac{1}{n \cdot (\lambda_4 - \lambda_3)} \cdot \sum_{j=1}^{\tilde{n}} m_j \cdot x_{1j}' \cdot \bar{y}_j = \frac{3}{45 \cdot (0{,}824 - 0{,}53)} \cdot \sum_{j=1}^{15} \left(x_{1j}^2 - \lambda_2\right) \cdot \bar{y}_j =$$
$$= 0{,}227 \cdot [(16 + 5 + 9 + 14 + 10 + 17 + 9) + 0{,}75 \cdot (14 + 8) -$$
$$-0{,}73 \cdot (11 + 8 + 16 + 16 + 17)] = -2{,}31;$$

$$b_{22} = 0{,}227 \cdot [0{,}27 \cdot (16 + 5 + 9 + 14 + 10 + 17 + 9) -$$
$$- 0{,}73 \cdot (14 + 8 + 16 + 16 + 17) + 0{,}75 \cdot (11 + 8)] = -3{,}32;$$

$$b_{33} = 0{,}227 \cdot [0{,}27 \cdot (16 + 5 + 9 + 14 + 10 + 17 + 9) -$$
$$- 0{,}73 \cdot (14 + 8 + 11 + 18 + 17) + 0{,}75 \cdot (16 + 16)] = 1{,}05.$$

Далее вычисляем

$$b_1 = \frac{1}{n \cdot \lambda_2} \cdot \sum m_j \cdot x_{1j} \cdot \bar{y}_j =$$
$$= 0{,}274 \cdot (-16 + 5 - 9 + 5 - 14 + 10 - 17 + 9 - 1{,}215 \cdot 14 + 1{,}215 \cdot 8) = -9{,}39;$$

$$b_2 = 0{,}274 \cdot (-16 - 5 + 9 + 5 - 14 - 10 + 17 + 9 - 1{,}215 \cdot 11 + 1{,}215 \cdot 8) = -2{,}39;$$
$$b_3 = 0{,}274 \cdot (-16 - 5 - 9 + 14 + 10 + 17 + 9 - 1{,}215 \cdot 16 + 1{,}215 \cdot 16) = 4{,}11;$$

$$b_{12} = \frac{1}{n \cdot \lambda_3} \cdot \sum_{j=1}^{15} m_j \cdot x_{1j} \cdot x_{2j} \cdot \bar{y}_j =$$
$$= \frac{3}{8} \cdot \sum_{j=1}^{15} x_{1j} \cdot x_{2j} \cdot \bar{y}_j = 0{,}375 \cdot (16 - 5 + 9 - 5 + 14 - 10 - 17 + 9) = 1{,}125;$$
$$b_{13} = 0{,}375 \cdot (16 - 5 + 9 - 5 + 14 - 10 - 17 + 9) = 1{,}125;$$
$$b_{23} = 0{,}375 \cdot (16 + 5 - 9 - 5 - 14 - 10 - 17 + 9) = 3{,}375.$$

Искомая модель имеет вид

$$\hat{y} = b_0' + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + b_{12} \cdot x_1x_2 + b_{13} \cdot x_1x_3 + b_{23} \cdot x_2x_3 +$$
$$+ b_{11} \cdot \left(x_1^2 - \lambda_2\right) + b_{22} \cdot \left(x_2^2 - \lambda_2\right) + b_{33} \cdot \left(x_3^2 - \lambda_2\right) =$$
$$= 11{,}67 - 9{,}39x_1 - 2{,}39x_2 + 4{,}11x_3 + 1{,}125x_1x_2 + 1{,}125x_1x_3 + 3{,}375x_2x_3 -$$
$$- 2{,}31 \cdot \left(x_1^2 - 0{,}73\right) - 3{,}32 \cdot \left(x_2^2 - 0{,}73\right) + 1{,}05 \cdot \left(x_3^2 - 0{,}73\right) = 15{,}01 - 9{,}39x_1 - 2{,}39x_2 +$$
$$+ 4{,}11x_3 + 1{,}215x_1x_2 + 1{,}215x_1x_3 + 3{,}375x_2x_3 - 2{,}31x_1^2 - 3{,}32x_2^2 + 1{,}05x_3^2.$$

5.5.1.2.3. Ротатабельные планы

Существенным недостатком ортогональных планов является то, что полученная на их основе модель с разной точностью предсказывает значения отклика в различных точках факторного пространства — т. е. дисперсии $S^2_{\hat{y}_j}$ неодинаковы для различных точек факторного пространства, равноудаленных от центра плана.

Этого недостатка лишены так называемые ротатабельные планы [621], которые позволяют получить модели с одинаковой дисперсией отклика во всех точках факторного пространства, равноудаленных от центра плана. Условие ротатабельности плана имеет вид

$$\lambda_4 = 3\lambda_3, \quad \text{или} \quad \sum_{j=1}^{n} x_{ij}^4 = 3 \sum_{j=1}^{n} x_{ij} x_{\varepsilon j}.$$

Формулы регрессионного анализа для симметричных планов второго порядка сохраняются и для ротатабельных планов с учетом условия ротатабельности. Свойство ротатабельности не зависит от числа n_0 опытов в центре плана (общее число опытов должно удовлетворять соотношению $\dfrac{\lambda_3}{\lambda_2^2} = \dfrac{k}{k+2}$ [621]).

Бокс и Хантер [627] предложили выбирать n_0, исходя из равномерности дисперсии предсказания. Такие планы называются униформ-ротатабельными и имеют место при значениях $\lambda_3^* = \dfrac{\lambda_3}{\lambda_2^2}$, удовлетворяющих условию

$$2\lambda_3^* \left(\lambda_3^* - 1\right)(k+2) + \lambda_3^* (k+1) - (k-1) = 0.$$

Например, если $k = 2, 3, 4, 5$, то λ_3^* должно соответственно равняться 0,7844; 0,8385; 0,8705; 0,8918. В этом случае отклик оценивается с примерно одинаковой дисперсией во всех точках шара с радиусом $\rho = \sqrt{\lambda_2}$.

Можно выбирать n_0, исходя из обеспечения ортогональности ротатабельного плана. Это условие соответствует требованию $\lambda_3^* = 1$.

Для композиционных планов условие ротатабельности обеспечивается соответствующим выбором звездного плеча $\alpha = 2^{\frac{k-p}{4}}$.

Число центральных опытов для композиционного ротатабельного ортогонального плана определяется соотношением

$$n_0 = \lambda_3^* = \left\{2^{k-p} + 42^{\frac{k-p}{2}} + 4\right\} - 2^{k-p} - 2k.$$

При соответствующих значениях λ_3^* можно найти n_0 для униформ- и ортогональных ротатабельных планов. Параметры и вспомогательные коэффициенты для некоторых ротатабельных планов приведены в табл. 245.

Таблица 245

Параметры униформ-ротатабельных планов

k	p	α^2	Распределение опытов			
			n	n_0	Вершины куба	Звездные точки
2	0	2	13	5	4	4
3	0	2,828	20	6	8	6
4	0	4	31	7	16	8
5	1	4	32	6	16	10

k	$n\lambda_2$	$n\lambda_3$	$n\lambda_4$	a	b	c	d	λ_3^*
2	8,000	4	12	2,600	1,300	1,625	−0,2438	0,81
3	13,657	8	24	3,327	1,136	1,250	−0,1378	0,86
4	24	16	48	4,429	1,107	0,969	−0,1153	0,86
5	24	16	48	5,091	1,091	1,000	−0,0900	0,89

5.5.1.2.4. D-оптимальные планы

Ортогональность и ротатабельность являются свойствами планов, а не критериями их оптимальности. Эти свойства не имеют количественных оценок.

Кифер [628] предложил выбирать в качестве критериев оптимальности величины, достигающие экстремума при выборе наилучшего способа обработки экспериментальных данных.

Такой величиной может быть объем эллипсоида рассеяния оценок параметров математической модели. Планы, минимизирующие объем эллипсоида рассеяния, т. е. планы, которые позволяют получить квадратичную модель с наиболее точными оценками коэффициентов, называются D-оптимальными. Различаются D-оптимальные планы с ограничениями на кубе $(-1 \leqslant x_i \leqslant +1)$ и на шаре $(\sum_{i=1}^{k} x_i^2 \leqslant 1)$.

Условия D-оптимальности плана на кубе имеют вид [629] $\lambda_3 = \dfrac{2k+1+\sqrt{4}k^2+12k+7}{4\,(k+2)}\lambda_2$, $(k > 1)$; $\lambda_4 = \lambda_2$.

Последнее равенство выполняется, если план эксперимента содержит только точки с координатами $-1, 0, +1$. Симметричные D-оптимальные планы построены Кифером (для $k \leqslant 5$) и Коно (для $k \leqslant 9$) [629]. Планы Кифера содержат 2^k точек в вершинах куба, $k \cdot 2^{k-1}$ точек в серединах ребер куба и $\dfrac{k\,(k-1)}{2}2^{k-2}$ точек

в центрах двумерных граней. Планы Коно включают в себя 2^k точек в вершинах куба, $k \cdot 2^{k-1}$ точек в серединах ребер куба и одну центральную точку. При $k = 2$ планы Кифера и Коно совпадают. D-оптимальность плана обеспечивается некоторым оптимальным распределением общего числа опытов по точкам плана, для точного соблюдения которого требуется большое число экспериментов. В связи с этим возникла задача построения планов, мало отличающихся от D-оптимальных и содержащих меньшее число опытов. Такие планы называются почти D-оптимальными и обозначаются B_i (i — число параметров модели) [630]. Матрицы планов B_4 и B_5 приведены в [630].

План второго порядка D-оптимален на шаре только тогда, когда он ротатабелен. Условия D-оптимальности на шаре имеют вид

$$\lambda_2 = \frac{k+3}{(k+1)(k+2)}; \quad \lambda_3 = \frac{k+3}{(k+1)(k+2)^2}; \quad \lambda_4 = 3\lambda_3 = \frac{3(k+3)}{(k+1)(k+2)^2}.$$

D-оптимальным планом на шаре будет композиционный план, включающий в себя 2^k точек в вершинах куба, вписанного в сферу единичного радиуса, 2^k точек на единичной сфере и одну центральную точку. При этом на каждую вершину куба должна приходиться доля наблюдений, равная $\dfrac{(k+3)\,k^2}{(k+1)(k+2)^2\,2^k}$, на каждую звездную точку — доля $\dfrac{k+3}{(k+1)(k+2)^2}$ и на центральную точку — доля $\dfrac{2}{(k+1)(k+2)}$. Практически обеспечение требуемого соотношения числа опытов в точках такого плана связано с необходимостью проведения большого количества экспериментов. Поэтому для обеспечения D-оптимальности на шаре обычно используется ротатабельное планирование, в котором число опытов в центре плана выбирается из условия D-оптимальности с отказом от требований ортогональности и униформности. Характеристики ротатабельных планов с числом центральных опытов, выбранным из условия D-оптимальности, приведены в табл. 246.

Таблица 246

Характеристики *D*-оптимального плана

k	P	α^2	Распределение опытов			
			n	n_0	Вершины куба	Звездные точки
2	0	2	11	3	4	4
3	0	8	16	2	8	6
4	0	4	26	2	16	8
5	1	4	27	1	16	10

k	$n\lambda_2$	$n\lambda_3$	$n\lambda_4$	a	b	c	d
2	8	4	12	3,667	1,833	1,3750	−0,5729
3	13,657	8	24	7,953	2,715	1,0000	−0,7270
4	24	16	48	13	3,25	0,8125	−0,6770
5	24	16	48	21	4,5	0,8438	−0,8438

5.5.1.2.5. Несимметричные планы второго порядка

Симметричные планы второго порядка позволяют бóльшую часть оценок коэффициентов регрессионной модели определить независимо друг от друга (коррелируются только оценки коэффициентов b_0 и b_{ii}, b_{ii} и b_{jj}, $i \neq j$). Отказ от требования

симметричности позволяет в некоторых случаях получить более эффективные планы. Наибольшее распространение получили несимметричные планы второго порядка, предложенные Хартли [631]. Планы Хартли экономичны, число экспериментов в них равно числу коэффициентов модели или ненамного превосходит его. Они близки к D-оптимальным, построены по композиционному принципу и состоят из дробной реплики ПФЭ, звездных точек и опытов в центре, число которых выбирается из условия D-оптимальности.

Планы Хартли нельзя сделать ортогональными или ротатабельными (кроме случая $k = 5$) выбором звездного плеча α. Если план Хартли образован на основе дробной реплики с генерирующим соотношением, не содержащим тройного взаимодействия, то он становится симметричным и регрессионный анализ функции отклика проводится по формулам симметричного планирования второго порядка.

Коэффициенты b_0, b_{ii}, а также b_i и $b_{\varepsilon\nu}$ для переменных x_i, x_ε и x_ν, не входящих в тройное взаимодействие генерирующего соотношения, подсчитываются по формулам

$$b_0 = \frac{a}{n}\sum_{j=1}^{n} y_j - \frac{b}{n}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 y_j; \quad b_{ii} = -\frac{b}{n}\sum_{j=1}^{n} y_j + \frac{c}{n}\sum_{j=1}^{n} x_{ij} y_j - \frac{d}{n}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 y_j;$$

$$b_i = \frac{1}{n\lambda_2}\sum_{j=1}^{n} x_{ij} y_j \ \ (i \neq 0); \quad b_{\varepsilon\nu} = \frac{1}{n\lambda_3}\sum_{j=1}^{n} x_{\nu j} x_{\varepsilon j} y_j \ \ (\varepsilon \neq \nu),$$

где a, b, c, d — коэффициенты, определенные ранее для симметричных планов.

Оценки коэффициентов $\tilde{b}_i$ и $\tilde{b}_{\varepsilon\nu}$, входящих в тройное взаимодействие определяющего контраста $J = x_i x_\varepsilon x_\nu$, вычисляются по формулам

$$\tilde{b}_i = \frac{e}{n}\left(\sum_{j=1}^{n} x_{ij} y_j - \sum_{j=1}^{n} x_{\varepsilon j} y_j\right); \quad \tilde{b}_{\varepsilon\nu} = -\frac{e}{n}\sum_{j=1}^{n} x_{ij} y_j + \frac{f}{n}\sum_{j=1}^{n} x_{ij} x_{\varepsilon j} x_{\nu j} y_j,$$

где $e = \dfrac{1}{\lambda_2 - \lambda_3}$; $f = \dfrac{\lambda_2}{\lambda_3} e = \dfrac{\lambda_2}{\lambda_3(\lambda_2 - \lambda_3)}$.

Дисперсии этих оценок равны соответственно $S^2_{\tilde{b}_i} = e\dfrac{S_y^2}{n}$ и $S^2_{\tilde{b}_{\varepsilon\nu}} = \dfrac{f}{n} S_y^2$. Дисперсии остальных коэффициентов модели вычисляются по формулам

$$S^2_{b_0} = \frac{a}{n} S_y^2; \quad S^2_{b_{ii}} = \frac{c - d}{n} S_y^2; \quad S^2_{b_i} = \frac{S_y^2}{n\lambda_2}; \quad S^2_{b_{ij}} = \frac{S_y^2}{n\lambda_3}.$$

Дисперсия предсказываемого значения отклика определяется формулой

$$S_{\hat{y}^2} = \frac{S_y^2}{n}\left[(c - d)\sum_{i=1}^{k} x_i^4 + \left(\frac{1}{\lambda_3} - 2d\right)\sum_{i<j}^{k} x_i^2 x_j^2 + \left(\frac{1}{\lambda_2} - 2b\right)\sum_{i=1}^{k} x_i^2 + a\right].$$

Параметры и вспомогательные коэффициенты некоторых планов Хартли для ограничений на кубе $(-1 \leqslant x_i \leqslant +1)$ и на шаре $(\sum_{i=1}^{k} x_i^2 \leqslant 1)$ приведены в табл. 247. Матрицы планов Хартли приведены в [630] для $k = 5$ и $k = 6$, там же приведены экономичные несимметричные планы Вестлейка ($k = 5$), Дрейпера, Хартли–Коно.

Таблица 247

Параметры планов Хартли

k	p	α^2	Генератор	Распределение опытов			
				n	n_0	Куб	Звезда
3	1	3	$x_1x_2x_3$	11	1	4	6
4	1	1	$x_1x_2x_3$	17	1	8	8
	1	4	$x_1x_2x_3$	17	1	8	8
5	1	1	$x_1x_2x_3x_4x_5$	27	1	16	10
	1	5	$x_1x_2x_3x_4x_5$	27	1	16	10

k	$n\lambda_2$	$n\lambda_3$	$n\lambda_4$	a	b	c	d	e	f	Тип плана
3	10	4	22	11	3,667	0,611	−1,14	1,833	4,5833	сфера
4	10	8	10	3,247	0,955	8,5	1,719	8,5	10,625	куб
	16	8	40	17	4,25	0,531	−0,996	2,125	4,25	сфера
5	18	16	18	3,727	0,818	13,5	2,455	–	–	куб
	28	16	66	27	5,4	0,54	−1,014	–	–	сфера

Задача 395. *Для плана Хартли на кубе для пяти переменных построить матрицу плана и по результатам его реализации найти поверхность отклика (в каждой точке плана проводились $m_j = 3$ измерения).*

Из табл. 247 находим для $k = 5$ (куб) $n = 27$, $n_0 = 1$, куб — 16, звезда — 10, т. е. всего проводятся 27 экспериментов, в том числе 1 — в центре плана, 16 — в точках куба, 10 — в звездных точках (звездное плечо $\alpha^2 = 1$, $\alpha = 1$).

Матрица плана и результаты эксперимента приведены в таблице:

j	x_1	x_2	x_3	x_4	x_5	$\bar{y}_j$	$\hat{y}_j$	j	x_1	x_2	x_3	x_4	x_5	$\bar{y}_j$	$\hat{y}_j$
1	+	+	+	+	+	1320	1288	15	+	+	−	+	−	1015	1002
2	−	−	+	+	+	257	265	16	−	−	−	+	−	202	230
3	−	+	−	−	−	434	443	17	0	0	0	0	0	427	380
4	+	−	−	−	−	465	478	18	+	0	0	0	0	511	590
5	−	+	−	+	+	451	445	19	−	0	0	0	0	293	266
6	+	−	−	+	+	467	465	20	0	+	0	0	0	498	591
7	+	+	+	−	−	1222	1205	21	0	−	0	0	0	312	230
8	−	−	+	−	−	257	280	22	0	0	+	0	0	475	513
9	−	+	+	+	−	511	509	23	0	0	−	0	0	379	353
10	+	−	+	+	−	493	494	24	0	0	0	+	0	412	431
11	+	+	−	−	+	407	386	25	0	0	0	−	0	337	330
12	−	−	−	−	+	183	202	26	0	0	0	0	+	329	381
13	−	+	+	−	+	471	460	27	0	0	0	0	−	508	468
14	+	−	+	−	+	442	435								

Вычисляем оценки коэффициентов регрессии (в нашем случае $n = \tilde{n} \cdot m_j = 27 \cdot 3 = 81$). Тогда имеем из табл. 247

$$\lambda_2 = \frac{18}{27} = 0{,}667; \quad \lambda_3 = \frac{16}{27} = 0{,}592; \quad \lambda_4 = \frac{8}{27} = 0{,}667;$$
$$a = 3{,}727; \quad b = 0{,}818; \quad c = 13{,}5; \quad d = 2{,}455.$$

Вычисляем:

$$b_0 = \frac{a}{n} \cdot \sum_{j=1}^{\tilde{n}} m_j \cdot \bar{y}_j - \frac{b}{n} \cdot \sum_{i=1}^{k} \sum_{j=1}^{\tilde{n}} x_{ij}^2 \cdot \bar{y}_j \cdot m_j = \frac{3{,}727}{81} \cdot \sum_{j=1}^{27} 3\bar{y}_j - \frac{0{,}818}{81} \cdot \sum_{i=1}^{5} \sum_{j=1}^{27} 3x_{ij}^2 \cdot \bar{y}_j = 380{,}1;$$

$$b_{ii} = -\frac{0{,}818}{81}\sum_{j=1}^{27}3\bar{y}_j + \frac{13{,}5}{81}\cdot\sum_{j=1}^{27}3x_{ij}\cdot\bar{y}_j - \frac{2{,}455}{81}\cdot\sum_{i=1}^{5}\sum_{j=1}^{27}3x_{ij}^2\cdot\bar{y}_j;$$

$$b_{11} = -\frac{0{,}818}{81}\sum_{j=1}^{27}3\bar{y}_j + \frac{13{,}5}{81}\cdot\sum_{j=1}^{27}3x_{1j}\cdot\bar{y}_j - \frac{2{,}455}{81}\cdot\sum_{i=1}^{5}\sum_{j=1}^{27}3x_{1j}^2\cdot\bar{y}_j = 27{,}9;$$

$$b_{22} = 30{,}9;\quad b_{33} = 52{,}9;\quad b_{44} = 0{,}4;\quad b_{55} = 44{,}4;$$

$$b_i = \frac{1}{18\cdot 3}\cdot\sum_{j=1}^{27}3x_{ij}\cdot\bar{y}_j = 0{,}0555\cdot\sum_{j=1}^{27}x_{ij\bar{y}_j};\quad b_1 = 0{,}0555\cdot\sum_{j=1}^{27}x_{1j}\cdot\bar{y}_j = 182{,}4;\quad b_2 = 180{,}6;$$

$$b_3 = 80{,}3;\quad b_4 = 50{,}6;\quad b_5 = 43{,}3;$$

$$b_{ij} = \frac{1}{16\cdot 3}\cdot\sum_{j=1}^{27}3x_{ij}\cdot x_{\varepsilon j}\cdot\bar{y}_j = 0{,}0625\cdot\sum_{j=1}^{27}x_{ij}\cdot x_{\varepsilon j}\cdot\bar{y}_j;\quad b_{12} = 0{,}0625\cdot\sum_{j=1}^{27}x_{1j}\cdot x_{2j}\cdot\bar{y}_j = 70{,}6;$$

$$b_{13} = 56{,}1;\quad b_{14} = 42{,}7;\quad b_{15} = -32{,}3;\quad b_{23} = 67{,}8;\quad b_{24} = 43{,}2;$$

$$b_{25} = -29{,}1;\quad b_{34} = -28{,}6;\quad b_5 = 38{,}4;\quad b_{45} = 71{,}8.$$

Искомая регрессионная модель имеет вид

$$\hat{y} = 380{,}1 + 182{,}4x_1 + 180{,}6x_2 + 80{,}3x_3 + 50{,}6x_4 + 43{,}3x_5 + 27{,}9x_1^2 + 30{,}9x_2^2 + \\ + 52{,}9x_3^2 + 0{,}4x_4^2 + 44{,}4x_5^2 + 70{,}6x_1x_2 + 56{,}1x_1x_3 + 42{,}7x_1x_4 - 32{,}3x_1x_5 + \\ + 67{,}8x_2x_3 + 43{,}2x_2x_4 - 29{,}1x_2x_5 - 28{,}6x_3x_4 + 38{,}4x_3x_5 + 71{,}8x_4x_5.$$

Вычисленные по полученной регрессионной модели значения отклика близки к экспериментальным реализациям наблюдаемой случайной величины.

5.5.2. Планирование экспериментов по поиску оптимума

Традиционные методы экспериментального нахождения оптимума функции нескольких переменных $y = f(x_1, x_2, \ldots, x_n)$ предполагают последовательное изучение зависимостей $y = f_i(x_i)$ от каждого из факторов при фиксированных значениях остальных. Недостатком таких методов является большое количество необходимых экспериментов и невозможность учесть взаимодействие факторов.

В последние годы разработаны специальные методы поиска оптимума, базирующиеся на математической теории экстремального эксперимента. Подробное изложение таких методов содержится в [631, 632].

5.5.2.1. Метод крутого восхождения

Бокс и Уилсон [626] предложили метод планирования экспериментов по поиску оптимума, сочетающий движение по градиенту функции отклика с ортогональным линейным планированием, — метод крутого восхождения. В соответствии с этим методом сначала проводится ПФЭ или ДФЭ с центром в некоторой точке факторного пространства с координатами $x_{10}, x_{20}, \ldots, x_{k0}$. Методами линейного регрессионного анализа вычисляются оценки коэффициентов модели отклика b_i $(i = 1, \ldots, k)$. Дальнейшее движение к оптимуму осуществляется изменением интервалов варьирования основных факторов в соответствии с уравнением

$$\tilde{x}_i - \tilde{x}_{i0} = \lambda b_i \Delta\tilde{x}_i,$$

причем используются некодированные значения переменных. Параметр λ выбирается следующим образом. Вычисляются произведения $b_i\Delta\tilde{x}_i$, и определяется базовый фактор x_δ, для которого $b_i\Delta\tilde{x}_\delta$ является наибольшим по абсолютной величине в сравнении с остальными факторами. Выбирается значение $\lambda = \lambda_1 = \dfrac{\varepsilon}{|b_\delta|}$,

где $0 < \varepsilon < 1$ (рекомендуется выбирать $\varepsilon = 0{,}5 \div 0{,}8$), b_ε — коэффициент линейной модели при факторе x_ε. Затем вычисляются шаги и координаты первой точки крутого восхождения по формуле $\tilde{x}_i^1 = \lambda_1 (b_i \Delta \tilde{x}_i) + \tilde{x}_{i0}$. Шаги и координаты последующих точек на линии крутого восхождения подсчитываются по формуле $\tilde{x}_i^j = j\lambda_1 (b_1 \Delta \tilde{x}_1) + \tilde{x}_{i0}$, где $j = 1, 2, \ldots$ — номер шага в направлении крутого восхождения.

Из всех реализованных опытов выбирается тот, который дал наилучшие результаты (например, максимальный отклик). Условия наилучшего опыта принимаются за основной уровень факторов, и цикл крутого восхождения повторяется до тех пор, пока все коэффициенты линейной модели b_i не будут признаны незначимыми. При движении к оптимуму значения параметра λ должны уменьшаться от цикла к циклу. Выбор величины шага при движении по градиенту производится из представлений исследователя о возможном характере оптимума.

При достижении оптимума для описания функции отклика используются методы планирования экстремальных экспериментов второго порядка. Движение по градиенту наиболее эффективно для симметричных регрессионных моделей, у которых коэффициенты b_i различаются несущественно.

Задача 396. *Необходимо найти оптимум отклика y в зависимости от трех факторов (x_1, x_2, x_3) методом крутого восхождения Бокса–Уилсона. Результаты последовательного экспериментирования по реализации крутого восхождения приведены в таблице*:

Реализация трехфакторного крутого восхождения

Номер по порядку	Последовательность операций крутого восхождения	Факторы			Отклик y_i
1	Основной уровень, $\tilde{x}_{i0}$	132	16	0,85	
2	Интервал варьирования, $\Delta\tilde{x}_i$	7	5	0,10	
3	Верхний уровень	139	21	0,95	
4	Нижний уровень	125	11	0,75	
5	Кодовое значение переменных	x_1	x_2	x_3	
6	Опыты ПФЭ 2^3				
	1	−	−	−	251
	2	+	−	+	303
	3	+	+	−	310
	4	−	−	+	324
	5	+	−	−	434
	6	−	+	−	470
	7	−	+	+	490
	8	+	+	+	1127
7	b_i	80	136	97	
8	$b_i\Delta\tilde{x}_i$	560	680	9,7	
9	$\lambda_1 = 0{,}8/\lvert b_\delta \rvert$		0,8/136		
10	Шаг $\approx \lambda_1(b_1\Delta\Delta\tilde{x}_i)$	3	4	0,06	
11					
12	Опыты на линии восхождения				
	1	135	20	0,91	1455
	2	138	24	0,97	1627
	3	141	28	1,03	1594

По результатам ПФЭ 2^3 находим коэффициенты регрессии при линейных параметрах:

$$b_1 = \frac{1}{n} \cdot \sum_{j=1}^{n} x_1 \cdot \bar{y}_j = 80; \quad b_2 = \frac{1}{n} \cdot \sum_{j=1}^{n} x_2 \cdot \bar{y}_j = 136; \quad b_3 = \frac{1}{n} \cdot \sum_{j=1}^{n} x_3 \cdot \bar{y}_j = 97.$$

Далее вычисляем произведения (указаны в таблице). Наибольшее значение произведения $b_2 \cdot \Delta\tilde{x}_2 = 680$. Поэтому выбираем фактор x_2 за базовый.

Выбираем для первого шага $\lambda = \lambda_1 = \dfrac{0,8}{|b_\delta|} = \dfrac{0,8}{136}$.

Тогда шаг в направлении крутого восхождения равен $\lambda_1 \cdot (b_1 \cdot \Delta\tilde{x}_1) = \dfrac{0{,}8 \cdot 80 \cdot 7}{136} = 3{,}0$; $\lambda_2 \cdot (b_2 \cdot \Delta\tilde{x}_2) = \dfrac{0{,}8 \cdot 136 \cdot 5}{136} = 4{,}0$; $\lambda_1 \cdot (b_3 \cdot \Delta\tilde{x}_3) = 0{,}06$.

Координаты точек крутого восхождения рассчитываем следующим образом: $\tilde{x}_1^1 = \lambda_1 \cdot (b_1 \cdot \Delta\tilde{x}_1) + \tilde{x}_{10} = 3 + 132 = 135$; $\tilde{x}_2^1 = 4 + 16 = 20$; $\tilde{x}_3^1 = 0{,}06 + 0{,}85 = 0{,}91$.

Второй шаг: $\tilde{x}_1^2 = 2\lambda_1 \cdot (\Delta\tilde{x}_1) + \tilde{x}_{i0} = 2 \cdot 3 + 132 = 138$, и т. д. Результаты опытов приведены в таблице, из которой видно, что второй опыт на линии крутого восхождения дает наибольший результат. Следовательно, максимум достигается при значениях координат точки $x_1 = 138$, $x_2 = 24$ и $x_3 = 0{,}97$. В окрестности этой точки можно теперь строить план второго порядка для уточнения поверхности отклика.

5.5.2.2. Симплексное планирование

Эффективным методом планирования эксперимента по поиску оптимума является последовательный симплексный метод, предложенный Спендли, Хексто и Химсуортом [634].

k-мерным симплексом называется выпуклый многогранник, образованный $(k+1)$ вершинами в k-мерном пространстве. Например, на плоскости ($k = 2$) симплексом будет треугольник. Симплекс называется правильным, или регулярным, если все его ребра равны между собой (например, равносторонний треугольник на плоскости). Из любого правильного симплекса можно, заменив одну точку на ее зеркальное отображение относительно границы или ребра симплекса, построить новый симплекс. Путем последовательной замены вершин можно осуществить перемещение симплекса в факторном пространстве.

Координаты вершин правильного k-мерного симплекса с единичным ребром и центром в начале координат определяются строками матрицы

$$\begin{vmatrix} -r_1 & -r_2 & -r_3 & \dots & -r_{k-1} & -r_k \\ R_1 & -r_2 & -r_3 & \dots & -r_{k-1} & -r_k \\ 0 & R_2 & -r_3 & \dots & -r_{k-1} & -r_k \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -R_{k-1} & -r_k \\ 0 & 0 & 0 & \dots & 0 & R_k \end{vmatrix},$$

где $r_i = \dfrac{1}{\sqrt{2i\,(i+1)}}$; $R_i = \dfrac{1}{\sqrt{i\,(2i+1)}}$.

Координата новой вершины симплекса после замены определяется формулой $x_{ik+2} = \dfrac{1}{k} \sum\limits_{j} x_{ij} - x'_i$, где x'_i — координата заменяемой вершины; $\dfrac{1}{k} \sum\limits_{j} x_{ij}$ — среднее значение координат всех точек симплекса, кроме заменяемой; x_{ik+2} — координата новой вершины.

Последовательное симплексное планирование эксперимента осуществляется следующим образом. Сначала ставятся эксперименты в вершинах симплекса (их число

на единицу больше числа факторов). Затем выявляется и отбрасывается вершина с минимальным (если необходимо найти максимум) или максимальным (если необходимо найти минимум) значением отклика. На оставшейся грани строится симплекс с новой вершиной, являющейся зеркальным отображением отброшенной. Затем проводится эксперимент в новой вершине симплекса, и цикл его перемещения повторяются до тех пор, пока симплекс не начнет „вращаться“ вокруг вершины с экстремальным значением отклика.

Симплексное планирование имеет ряд существенных достоинств — оно обладает свойством самоконтроля, исключает влияние ошибок эксперимента на конечный результат оптимизации; на любом этапе экспериментирования можно включить в рассмотрение еще один фактор, добавив только одну вершину (экспериментальную точку), увеличивающую размерность симплекса на единицу.

Метод не предъявляет жестких требований к точности фиксирования значений отклика (он требует только ранжирования их по величине). Отсутствует необходимость описания поверхности отклика. Симплекс-планирование безразлично к форме поверхности отклика. Подробно метод симплекс-планирования изложен в [621].

Задача 397. *Рассмотрим пример поиска оптимума симплекс-планированием. Предположим, имеются два фактора x_1 и x_2 с основными уровнями $\tilde{x}_{10} = 1000$ и $\tilde{x}_{20} = 1$ и интервалы варьирования $\Delta\tilde{x}_1 = 500$ и $\Delta\tilde{x}_2 = 0{,}2$ соответственно.*

Симплексом в нашем случае является треугольник ABC. Будем ставить эксперимент в точке A с координатами в кодированной форме

$$x_1 = \frac{\tilde{x}_1 - x_{10}}{\Delta x_{10}} \quad \text{и} \quad x_2 = \frac{\tilde{x}_2 - x_{20}}{\Delta x_{20}}.$$

Затем ставим эксперименты в точках B с координатами $x_1 = 0{,}5$ и $x_2 = 0{,}86$ и в точке C с координатами $x_1 = 1$ и $x_2 = 0$.

Дальнейшее движение симплекса приведено в таблице:

Номер опыта, j	Симплекс	Вершина, в которой ставится опыт	x_1		x_2		Отклик, y_j
			Код	Значение	Код	Значение	
1	ABC	A	0	1000	0	1	9
2	ABC	B	0,5	1250	0,86	1,172	13
3	ABC	C	1	1500	0	1	11
4	BCA'	A'	1,5	1750	0,86	1,172	17
5	$BC'A'$	C'	1	1500	1,72	1,344	18
6	$B'C'A'$	B'	2	2000	1,72	1,344	21
7	$B'C'A''$	A''	1,5	1750	2,58	1,516	24
8	$B'C''A''$	C''	2,5	2250	2,58	1,516	20

Видим, что из точек стартового комплекса ABC наименьший отклик имеется в точке A ($y_j = 9$). Так как ищется максимум, то будем искать точку A', симметричную к точке A относительно грани BC симплекса. Имеем координаты искомой точки

$$x_{14} = \frac{2}{2}\cdot(0{,}5+1) - 0 = 1{,}5; \quad x_{24} = \frac{2}{2}\cdot(0{,}86+0) - 0 = 0{,}86.$$

Ее натуральные (некодированные) координаты равны $x_{14} = \dfrac{\tilde{x}_1 - \tilde{x}_{10}}{\Delta\tilde{x}_1}$ и $\tilde{x}_1 = \Delta\tilde{x}_1 \times$ $\times\, \tilde{x}_{14} = 500\cdot 1{,}5 + 1000 = 1750$; $\tilde{x}_2 = \Delta\tilde{x}_2\cdot\tilde{x}_{24} + \tilde{x}_{20} = 0{,}86\cdot 0{,}2 + 1 = 1{,}172$.

Далее по аналогии рассчитываются остальные точки симплекс-планирования (результаты последовательного симплекс-планирования приведены в таблице).

Видим из таблицы, что наибольшее значение достигается в точке A'' с координатами $x_1 = 1{,}5$; $x_2 = 2{,}58$.

Очень короткое послесловие

Уважаемый читатель! Лежащая перед тобой книга — не классическое исследование, по окончании которого уместно делать вывод о том новом, что́ автору удалось вырвать у природы. Эта книга — и не учебник, призванный системно ввести тебя, мой читатель, в предмет. Скорее всего, эта книга — справочник с комментариями, заменяющими изложение основ теории обработки результатов наблюдений.

Однако пространные комментарии, большое число и подробное описание примеров позволяют автору надеяться на то, что его труд может претендовать на статус пособия или руководства, т. е. быть чем-то средним между классическим учебником и ординарным справочником.

Автор хотел бы ограничиться одной ремаркой, цель которой — ответить на вероятный вопрос оппонента-профессионала: почему пропущены (или недостаточно полно изложены) некоторые разделы прикладной математической статистики? Такие, например, как многомерная статистика, временны́е ряды, методы распознавания незрительных образов, теория обслуживания и очередей и т. п.

Вопрос естествен. Однако автор хотел бы обратить внимание оппонента на вводную часть книги. Цель книги — дать инструмент в руки инженера или исследователя, не имеющего горячего стремления погружаться в математические глубины теории вероятностей и математической статистики. Автор исходил из того, что математика перечисленных выше разделов прикладной статистики весьма „тяжела“ для среднего инженера или исследователя и вряд ли будет им воспринята. Впрочем, может быть, это слабый аргумент, и дело скорее в неспособности автора соответствующим образом изложить эти разделы прикладной статистики.

Так или иначе, автор сделал всё, что мог, и надеется, что тем самым внес свой скромный вклад во введение инженеров и исследователей в увлекательный мир обработки результатов наблюдений.

Последние строки книги автор считает своим долгом посвятить словам благодарности ее научному редактору В. С. Ароловичу, чей напряженный труд немало способствовал превращению исходного авторского текста в книгу, достойную внимания читателя.

Список литературы

1. *Крамер Г.* Математические методы статистики / Пер. с англ. — М.: Мир, 1975.
2. *Ван дер Варден Б. Л.* Математическая статистика / Пер. с нем. — М.: ИЛ, 1960.
3. *Колмогоров А. Н.* Основные понятия теории вероятностей. — М.: Наука, 1974.
4. *Гнеденко Б. В.* Курс теории вероятностей. — М.: Наука, 1965.
5. *Кендалл М.*, *Стьюарт А.* Теория распределений / Пер. с англ. — М.: Наука, 1966.
6. *Уилкс С. С.* Математическая статистика / Пер. с англ. — М.: Наука, 1967.
7. *Юл. Д. Э.*, *Кендалл М. Д.* Теория статистики / Пер. с англ. — М.: Госстатиздат, 1960.
8. *Джини К.* Средние величины / Пер. с итал. — М.: Статистика, 1970.
9. *Закс Л.* Статистическое оценивание / Пер. с нем. — М.: Статистика, 1976.
10. *Митропольский А. К.* Техника статистических вычислений. — М.: Наука, 1971.
11. *Идье В.*, *Драйад Д.*, *Джеймс Ф.*, *Рус М.*, *Садуле Б.* Статистические методы в экспериментальной физике. — М.: Атомиздат, 1976.
12. *Смирнов Н. В.*, *Дунин-Барковский И. В.* Курс теории вероятностей и математической статистики. — М.: Наука, 1969.
13. *Хан Г.*, *Шапиро С.* Статистические модели в инженерных задачах / Пер. с англ. — М.: Мир, 1969.
14. *Химмельблау Д.* Анализ процессов статистическими методами / Пер. с англ. — М.: Мир, 1970.
15. *Шторм Р.* Теория вероятностей. Математическая статистика. Статистический контроль качества / Пер. с нем. — М.: Мир, 1970.
16. Справочник по надежности. Т. 1 / Пер. с англ. — М.: Мир, 1969.
17. *Хастингс Н.*, *Пикок Дж.* Справочник по статистическим распределениям / Пер. с англ. — М.: Статистика, 1980.
18. *Холлендер М.*, *Вулф Д.* Непараметрические методы статистики / Пер. с англ. — М.: Финансы и статистика, 1983.
19. *Йэйтс Ф.* Выборочный метод в переписях и обследованиях. — М.: Статистика, 1965.
20. *Пустыльник Е. И.* Статистические методы анализа и обработки результатов наблюдений. — М.: Наука, 1968.
21. *Мот Ж.* Статистические предвидения и решения на предприятии. — М.: Прогресс, 1966.
22. *Коуден Д.* Статистические методы контроля качества / Пер. с англ. — М.: Наука, 1961.
23. *Абезгауз Г. Г. и др.* Справочник по вероятностным расчетам. — М.: Воениздат, 1970.
24. *Оуэн Д. Н.* Сборник статистических таблиц / Пер. с англ. — М.: ВЦ АН СССР, 1966.
25. *Большев Л. Н.*, *Смирнов Н. В.* Таблицы математической статистики. — М.: Наука, 1965.
26. *Келли Т. Л.* Статистические таблицы / Пер. с англ. — М.: ВЦ АН СССР, 1966.
27. *Митропольский А. К.* Интеграл вероятностей. Л.: Изд-во ЛГУ, 1972.
28. Таблицы нормального интеграла вероятностей, нормальной плотности и ее нормированных производных. — М.: Изд-во АН СССР, 1960.
29. *Янко Я.* Математико-статистические таблицы / Пер. с чеш. — М.: Госстатиздат, 1961.
30. *Дьяконов В. П.* Справочник по расчетам на микрокалькуляторах. — М.: Наука, 1986.
31. *Цветков А. Н.*, *Епанечников В. А.* Прикладные программы для микро-ЭВМ „Электроника Б3-34“, „Электроника-МК-56“, „Электроника МК-54“. — М.: Финансы и статистика, 1984.
32. Справочник по специальным функциям / Под ред. М. Абрамовица и И. М. Стиган. — М.: Наука, 1979.
33. *Епанечников В. А.*, *Цветков А. Н.* Справочник по прикладным программам для микрокалькуляторов. — М.: Финансы и статистика, 1980.

34. *Градштейн И. С., Рыжик И. М.* Таблицы интегралов, сумм, рядов и произведений. — М.: Наука, 1971.
35. *Hastings C.* Approximations for digital computers, Princeton Univ. Prtss., 1955.
36. *Казакавичюс К. А.* Приближенные формулы для статистической обработки результатов механических испытаний // Завод. лаб. 1988. Т. 54, № 12. С. 82–85.
37. Надежность и эффективность в технике. Справочник / Под ред. Б. В. Гнеденко. — Т. 2. — М.: Машиностроение, 1987.
38. *Kennedy W. J., Gentle J. E.* Statistical computing, Marcel Dreker Inc., New York and Basel, 1980.
39. *Joiner B. I., Rosenblatt J. R.* Some properties of the in samples from Tukey's symmetric lambda distribution // JASA. 1971. V. 66. P. 394–399.
40. *Lin Jinn-Tyan.* Approximating the normal tail probability and its inverse for use on a pocket calculator // JRSS. Sec. C. 1989. V. 38, № 1. P. 69–70.
41. *Goldwaite L. R.* Failure rate study for the lognormal lifetime model // Proc. Symp. Reliability Quality Control, Philadelphia, 1961.
42. *Aitchison J., Brown J.* The lognormal distribution. London: Cambridge Univ. Press, 1951.
43. *Венецкий И. Г., Венецкая В. И.* Основные математико-статистические понятия и формулы в экономическом анализе. — М.: Статистика, 1982.
44. *Шор Я. Б., Кузьмин Ф. И.* Статистические методы анализа и контроля качества и надежности. — М.: Сов. радио, 1968.
45. *Пагурова В. И.* Таблицы неполной гамма-функции. — М.: ВЦ АН СССР, 1963.
46. *Шор Я. Б.* Статистические методы анализа и контроля качества и надежности. — М.: Сов. радио, 1962.
47. *Слуцкий Е. Е.* Таблицы для вычисления неполной Г-функции и функции вероятностей. — М.: Изд-во АН СССР, 1950.
48. *Лившиц И. Г.* Обобщенное гамма-распределение // Надежность и контроль качества. 1975. № 7. С. 28–34.
49. *Stacy E. W.* A generalization of the gamma-distribution // AMS. 1962. V. 28. P. 1187–1192.
50. *Пирсон К.* Таблицы неполной бета-фукнции. — М.: ВЦ АН СССР, 1974.
51. *Хальд А.* Математическая статистика с техническими приложениями. — М.: ИЛ, 1956.
52. *Cadwell L. H.* An approximation to the symmetrical incomplete beta function // Biometrika. 1952. V. 39. P. 204–207.
53. *Wise M. E.* The incomplete beta-function, as a contour integral and a quickly conversing series for its inverse // Biometrika. V. 37. 1950. P. 208–218.
54. *Wise M. E.* On normalizing the incomplete beta-function for fitting to dosage response curves // Biometrika. 1969. V. 47. P. 173–175.
55. *Molenar W.* Approximations to the Poisson, binomial and hypergeometric distribution function, Math. Center tracts., 31, Amsterdam, 1970.
56. *Мюллер П., Нойман П., Шторм Р.* Таблицы по математической статистике. — М.: Финансы и статистика, 1982.
57. *Ликеш И., Ляга Й.* Основные таблицы математической статистики. — М.: Финансы и статистика, 1985.
58. *Boyd W. C.* A nomogram for chi-square // JASA. 1965. V. 60. № 309. P. 344–346.
59. *Fisher R. A., Yates F.* Statistical tables for biological, agriculture and medical research. Edinburgh & London: Oliver and Boyd, 1963.
60. *Fisher R. A.* Statistical tables for research workers. Edinburgh: Oliver and Boyd, 1925.
61. *Wilson T. B., Hilferty M. M.* The distribution of chi-square // Proc. of the national Academy of sciences. 1931. V. 17, № 12. P. 684–688.
62. *Hawkins D. M., Wixley R. A. J.* A note on the transformation of the normality // AS. 1986. V. 40, № 4. P. 296–298.
63. *Hernandes F., Johnson R. A.* The large-sample behavior of transformations to normality // JASA. 1980. V. 75. P. 855–861.
64. *Taylor J. M. G.* Power transformations to symmetry // Biometrika. 1985. V. 72. P. 145–152.
65. *Goldstein R. B.* Chi-square quantiles, Algorithm 51 // Commun. Assoc. Comp. 1973. V. 16. P. 483–485.

66. *Severo N. C., Zelen V.* Normal approximation to the chi-square and non-central F-probability function // Biometrika. 1960. V. 47. P. 411–416.
67. *Haldane J. B. S.* The approximate normalization of the class of frequency distribution // Biometrika. V. 29, 1937. P. 392–404.
68. *Golberg H., Levine H.* Approximate formulas for the percentage points and normalization of t and χ^2 // AMS. 1945. V. 17. P. 216–225.
69. *Gilbert R. J.* A sample formula for cuterpolating tables of χ^2 // Biometrics. 1977. V. 33. P. 383–385.
70. *Hoagling D. C.* Direct approximations for chi-square percentage points // JASA. 1977. V. 72, № 359. P. 508–515.
71. *Aroian L. A.* A new approximation to the levels of significance of the chi-square distribution // AMS. 1973. V. 14. P. 93–95.
72. *Peizer D. B., Pratt J. W.* A normal approximation for binomial, F, beta and other common, related tail probability, 1 // JASA. 1968. V. 63, № 324. P. 1416–1456.
73. *Pratt J. W.* A normal approximate for binomial, F, beta and other common, related tail probability, 2 // JASA. 1968. V. 63, № 324. P. 1457–1483.
74. *Ling R. F.* A study of the accuracy of same approximations for t, χ^2 and F tail probability // JASA. 1978. V. 73, № 362. P. 274–283.
75. *Zar J. H.* Approximations for the percentage points of the chi-square distribution // JRSS. Sec. C. 1978. V. 27, № 3. P. 280–290.
76. *Fisher R. A.* Expansion of „Student's“ integral in power of n^{-1} // Metron. 1926. № 5. P. 109–112.
77. *Koechler K. J.* A simple approximations for the percentage of t distribution, // Technometrics. 1983. V. 25, № 1. P. 103–105.
78. *Koechler K. J.* A simple approximations for the percentage of t distribution // JRSS. Sec. C. 1978. V. 27, № 3. P. 280–290.
79. *Nelson L. S.* Same notes of Student's // J. of Quality Techn. 1984. V. 16, № 1. P. 64–65.
80. *Nelson L. S.* Standardizing Student's t // AS. 1973. V. 27. P. 93.
81. *Gentleman W. M., Jenkins M. A.* An approximation for Student's t-distribution // Biometrika. 1968. V. 55. P. 571–572.
82. *Scott A., Smith T. M. E.* A note on Moran's approximation to Student's // Biometrika. 1970. V. 57, № 3. P. 681–682.
83. *Dawson F. H.* Alternatives to the use of tabulated values of distribution in statistical programs // Nature. 1975. V. 256. P. 148.
84. *Moran P. A. P.* Accurate approximation for t-test, in Festschrift for J. Nayman: Research papers in statistics, ed. F. N. David. — N. Y.: Wiley, 1966. P. 225–230.
85. *Wang Y. Y.* Probability the type 1 errors of the Welch tests for the Behrens–Fisher problem // JASA. 1971. V. 66. P. 605–608.
86. *Fenstad G. U.* A comparison between the U and V tests in the Behrens–Fisher problem // Biometrika. 1983. V. 70, № 1. P. 300–302.
87. *Мардиа К., Земроч П.* Таблицы F-распределений и распределений, связанных с ними. — М.: Наука, 1984.
88. *Paulson E.* An approximate normalization of the analysis of variance distribution // AMS. 1942. V. 13. P. 233–235.
89. *Haines P. D.* A closed form approximation for calculating the percentage points of the F and t distributions // JRSS. Sec. C. 1988. V. 37, № 1. P. 95–100.
90. *Carter A. H.* Approximation to percentage points of the z-distribution // Biometrika. 1947. V. 34. P. 352–358.
91. *Hald A. D.* Statistical tables and formulas, New York, 1952. P. 47–59.
92. *Cochran W. G.* Note on an approximate formula for the significance levels of z // AMS. 1940. V. 11. P. 93–95.
93. *Бородачев Н. А., Абдрашитов Р. М., Веселова И. М.* Точность производства в машиностроении и приборостроении. — М.: Машиностроение, 1973.
94. *Гумбель Э.* Статистика экстремальных значений. — М.: Мир, 1965.

95. *Капур К., Ламберсон Л.* Надежность и проектирование в технике. — М.: Мир, 1980.
96. *Owen D. B.* The power of Student's t-test // JASA. 1965. V. 60, №309. P. 320–333.
97. *Джонсон Н., Лион Ф.* Статистика и планирование эксперимента в технике и науке. Методы обработки данных. — М.: Мир, 1980.
98. *Pearson E. S.* Note on an approximation to the distribution of non-central χ^2 // Biometrika. 1959. V. 46. P. 364–366.
99. *Patnaik E. S.* The non-central χ^2 and F distribution and their applications // Biometrika. 1949. V. 36. P. 202–232.
100. *Pearson E. S., Hartley H. O.* Charts of the power function for analysis of variance test, derived from the non-central F-distribution // Biometrika. 1951. V. 38. P. 112–130.
101. *Tiku M. L.* Laguerre series forms of non-central χ^2 and F distribution // Biometrika. 1965. V. 52. P. 415–427.
102. *Tiku M. L.* A note on approximating to the non-central F distribution // Biometrika. 1966. V. 53. P. 603–610.
103. *Судаков Р. С. и др.* Статистические задачи обработки систем и таблицы для числовых расчетов показателей надежности. — М.: Высшая школа, 1975.
104. *Большев Л. Н.* Асимптотические пирсоновские преобразования // Теор. вероятн. и ее примен. 1963. №8. С. 129–155.
105. *Браунли К. А.* Статистическая теория и методология в науке и технике. — М.: Наука, 1977.
106. *Рао С. Р.* Линейные статистические методы и их приложения. — М.: Наука, 1968.
107. *Anscomb F. J.* The transformation of Poisson, binomial and negative-binomial data // Biometrika. 1948. V. 35. P. 246–254.
108. *Черницкий П. И.* Таблицы вероятностей. — М.: Воениздат, 1957.
109. *Binns M. R.* Approximating the negative binomial via the positive binomial // Technometrics. 1974. V. 16. P. 323–324.
110. Best D. J., Gipps P. G. An improved gamma approximation to the negative binomial // Technometrics. 1974. V. 16. P. 621–624.
111. *Liberman G. J., Owen D. B.* Tables of the hypergeometric probability distribution. Stanford Univ. Press, Stanford, Calif., 1961.
112. *Vilaplana J. H.* A tables of the hypergeometric probability distribution, Publ. Lab. Ing. Ind. Panama, s. B, r. 2, 1976, P. 1–2095.
113. *Беляев Ю. К.* Вероятностные методы выборочного контроля. — М.: Наука, 1975.
114. *Локтев А. Л.* Аппроксимационные формулы для распределений нормального, Стьюдента, хи-квадрат // Надежность и контроль качества. 1990. №5. С. 22–25.
115. *Редько М. Ю.* Об аппроксимациях нецентральных F-распределений центральным // Надежность и контроль качества. 1990. №5. С. 52–59.
116. *Вульфович Б. А.* Оценивание параметров малой выборки. — Деп. во ВНИЭРХ 22.10.91, №1180–рх9.
117. *Dixon W. J.* Estimation of the mean and standard deviation of a normal population // AMS. V. 28. P. 806–809.
118. *Боярский Э. А.* Порядковые статистики. — М.: Статистика, 1972.
119. Введение в теорию порядковых статистик / Пер. с англ. — Под ред. А. Сархана и Б. Гринберга. — М.: Статистика, 1970.
120. *Pearson E. S., Tukey J. W.* Approximable means and standard deviations based on distances between persenage points of frequency // Biometrika. 1965. V. 52. P. 533–546.
121. *Кенуй М. Г.* Быстрые статистические вычисления. Упрощенные методы оценивания и проверки. — М.: Статистика, 1979.
122. *Дэйвид Г.* Порядковые статистики / Пер. с англ. — М.: Наука, 1979.
123. *Eisenberger I., Posner E. S.* Systematic statistics used for data compression in space telemetry // JASA. 1965. V. 60. P. 97–133.
124. *Walsh J. E.* Some significance tests for the median, which are valid under very general conditions // AMS. 1949. V. 20. P. 64–81.
125. *Ашмарин И. П., Васильев Н. Н., Амбросов В. А.* Быстрые методы статистической обработки и планирования экспериментов. — Л.: Изд-во ЛГУ, 1971.

126. *Фишбейн В. И.* Номограмма, реализующая функцию биномиального распределения // Надежность и контроль качества. 1972. № 12. С. 51–63.
127. *Herrey E. M. J.* Confidence intervals based on the mean absolute deviation of normal sample // JASA. 1965. V. 60. P. 257–269.
128. *Krutchkoff F. G.* The correct use of the sample mean absolute deviation in confidence intervals for a normal variate // Technometrics. 1966. V. 8. P. 663–67.
129. *Peters C. A. F.* Über die Bestimmung des Wahrscheinlichen Fehlers einer Beobachtung aus den Abweichungen der Beobachtungen von ihrem arithmetischen Mittel // Astronomische Nachrichten. 1956. 44. S. 30–31.
130. *Gutterman H. E.* An upper bound for the sample standard deviation // Technometrics. 1962. V. 4. P. 134–135.
131. *Dawton F.* Linear estimates with polynomial coefficients // Biometrika. 1966. V. 53. P. 129–141.
132. *Кендалл М. Дж., Стьюарт А.* Статистические выводы и связи. — М.: Наука, 1973.
133. *Зелингер Г. Я.* Построение кратчайших доверительных интервалов для среднеквадратического отклонения нормального распределения // Завод. лаб. 1983. № 3. С. 64–66.
134. *Iliescu D. V., Voda V. Ch.* On the estimation of standard deviation for a normal population, Trab. estadist, J. Invest. Oper. 1974. V. 25. P. 71–98.
135. *Cohen A. C.* Tables for maximum likelihood estimates-singly truncated and singly censored samples // Technometrics. 1961. V. 3. P. 535–541.
136. *Schmee J., Gladstein D., Nelson W.* Confidence limits for parameters of a normal distribution from singly censored samples, using maximum likelihood // Technometrics. 1985. V. 27. P. 119–128.
137. *Gupta A. K.* Estimation of the mean and standard deviation of a normal population from a censored sample // Biometrika. 1952. V. 39. P. 260–273.
138. *Sarhan A. E., Greenberg B. G.* Estimation of location and scale parameters by order censored sample, part 3, Techn. Report, 4, OOR Project, 1957.
139. *Dixon W. J.* Simplified estimation from censored normal samples // AMS. 1954. V. 25. P. 610–614.
140. *Гнеденко Б. В., Беляев Ю. К., Соловьев А. Д.* Математические методы в теории надежности. — М.: Наука, 1965.
141. *Pandley B. N., Srivostava R.* A shrinkage estimator for scale parameter of an exponential distribution // Microelectron Reliability. 1987. V. 27. P. 949–951.
142. *Harter H. L.* Estimating the parameters of negative exponential populations from one or two order statistics // AMS. 1961. V. 32. P. 1078–1090.
143. *Epstein B.* Estimation from life test data // Technometrics. 1960. V. 2. P. 447–454.
144. *Ogawa J.* Determination of optimal spacing for the estimation of the side parameters of an exponential distribution based on sample quantiles // Ann. Inst. St. Mat. 1960. V. 12. P. 11–155.
145. *Harber H. L.* Best linear unbiased under symmetric censored of the parameters of a double exponential population // JASA. 1961. V. 32. P. 1078–1084.
146. *Kuldorff G.* Estimation of one or two parameters of exponential distribution on the basis of suitably chosen order statistics // AMS. 1963. V. 34. P. 1419–1431.
147. *Siddiqui M. M.* Optimal estimators of the parameters of negative exponential distribution from one or two order statistics // AMS. 1963. V. 34. P. 117–121.
148. *Saleh A. K., Ali M. M.* Asimptotic optimum quantiles for the estimation of the parameters of the negative exponential distribution // AMS. 1966. V. 37. P. 143–151.
149. *Hassanein K. M.* Estimation of the parameters of the extreme values distribution by use of two or three order statistics // Biometrika. 1969. V. 56. P. 429–436.
150. *Weissman I.* Estimation of parameters and large quantiles based on the K largest observations // JASA. 1978. V. 73. P. 812–815.
151. *Kubat P., Epstein B.* Estimation of quantiles of location-scale distribution based on two or three order statistics // Technometrics. 1980. V. 22. P. 575–581.
152. *Кудлаев Э. М.* Оценивание параметров распределения Вейбулла–Гнеденко (обзор) // Изв. АН СССР: Техн. кибернет. 1986. № 6. С. 5–18.

153. *Мартыненко Ю. Н.* Ускоренная оценка параметров распределения Вейбулла методом максимального правдоподобия // Надежность и контроль качества. 1987. № 6. С. 11–13.
154. *Подольский Ю. М.* О точном вычислении некоторых решеточных функций Грина // Укр. физ. ж. 1972. Т. 17, № 2. С. 240–260.
155. *Tonaka S., Ichiakawa M.* Approximate formula of coefficient of variation for Weibull distribution // Reliability Eng. 1983. V. 4. P. 141–143.
156. *Глушко В. Т., Бобро Н. Т., Рубец Г. Т., Гажемон Л. И.* Некоторые методы оценки параметров обобщенного распределения Вейбулла // Надежность сложных технических систем: Сб. — Киев: Наукова думка, 1974. С. 83–91.
157. *Один И. М.* Определение параметров распределения Вейбулла методом наименьших квадратов // Надежность и контроль качества. 1975. № 7. С. 45–48.
158. *Багно А. Н.* Об одном методе оценки параметров распределения Вейбулла // Надежность и контроль качества. 1975. № 7. С. 25–32.
159. *Балицкая Е. О., Золотухина Л. А.* Асимптотическая эффективность моментных и квантильных оценок параметров распределения Вейбулла // Завод. лаб. 1988. Т. 54, № 3. С. 92–96.
160. *Bain L. J.* Inferences based on censored sampling from the Weibull or extreme-value distribution // Technometrics. 1972. V. 14. P. 693–702.
161. *Engelhart M.* On simple estimation of the parameters of the Weibull or extreme-value distribution // Technometrics. 1975. V. 17. P. 369–373.
162. *Mann N. R., Fertig K. W.* Simplified efficient point and interval estimators for Weibull parameters // Technometrics. 1975. V. 17. P. 361–368.
163. *Engelhart M., Bain L. J.* Some complete and censored sampling results for the Weibull or extreme-value distribution // Technometrics. 1973. V. 15. P. 541–549.
164. *Engelhart M., Bain L. J.* Some results on point estimators for the two-parameter Weibull or extreme-value distribution // Technometrics. 1974. V. 16. P. 49–56.
165. *Mann N. R.* Best linear invariant estimation for Weibull parameters under progressive censoring // Technometrics. 1971. V. 13. P. 521–533.
166. *Балицкая Е. О., Золотухина Л. А.* Оценка параметров распределения Релея // Завод. лаб. 1992. Т. 58, № 5. С. 54–57.
167. *Mann N. R., Fertig K. W.* Tables for obtaining Weibull confidence bounds and tolerance bounds on best linear invariant estimates of parameters of the extreme-value distribution // Technometrics. 1973. V. 15. P. 87–101.
168. *Stasy E. W.* Quasymaximum likelihood estimations for two-parameter gamma-distribution // IBM J. Res. And Develop. 1973. V. 17. P. 115–124.
169. *Shenton L. R., Bowman K. O.* Further remarks on maximum likelihood estimator for gamma-distribution // Technometrics. 1972. V. 14. P. 725–733.
170. *Рудых Г. А.* Оценки максимального правдоподобия двухпараметрического гамма-распределения // Надежность и контроль качества. 1977. № 1. С. 60–65.
171. *Мартыненко Ю. Н.* Ускоренная оценка параметров гамма-распределения // Надежность и контроль качества. 1984. № 12. С. 10–11.
172. *Буймов А. Г., Буймова Н. А.* Об имитации и оценивании параметров гамма-распределения // Исследование корреляционно-экстремальных систем, Томск, 1987, С. 8–11.
173. *Wang Wenjun.* Scale parameter of gamma-distribution and its autocovariance estimation // Chin. J. Appl. Prob. and Stat. 1987. V. 3. P. 193–202.
174. *Clopper C. I., Pearson E. S.* The use of confidence or fiducial limits illustrated in the case of the binomial // Biometrika. 1934. V. 26. P. 404–413.
175. *Чурилов В. И.* Об использовании приближенных зависимостей для определения доверительных границ параметра биномиального распределения // Надежность и контроль качества. 1972. № 12. С. 51–63.
176. *Blyth C. R.* Approximate binomial confidence limits // JASA. V. 81, 1986. P. 843–855.
177. *Hall P.* Improving the normal approximation when constructing one-sided confidence intervals for binomial and Poisson parameters // Biometrika. 1982. V. 69. P. 647–652.

178. *Paulson E.* An approximate normalization of the analysis of variance distribution // AMS. 1942. V. 13. P. 233–235.
179. *Camp B. H.* Approximation to the point binomial // AMS. 1951. V. 22. P. 130–131.
180. *Pratt J. W.* A normal approximation for binomial, F, beta, and other common related tail probabities. 2 // JASA. 1968. V. 63. P. 1457–1483.
181. *Титенко И. М.* Оценка доверительных границ параметра биномиального распределения // Надежность и контроль качества. 1983. №12. С. 31–36.
182. *Дамидович Н. О., Курский И. Ю. и др.* О приближенных формулах вычисления доверительных границ вероятностей при биномиальном плане испытаний // Завод. лаб. 1988. Т. 54, №3. С. 95–98.
183. *Noether G. E.* Some simple distribution free confidence intervals for the center of a symmetric distribution // JASA. 1973. V. 68. P. 716–719.
184. *Багдонавичюс В., Адоменас В.* Исследование одного распределения и оценок его параметров // Применение теории вероятностей и математической статистики: Сб. Вып. 5. Вильнюс, 1983. — С. 123–130.
185. *Hahn G. J.* A simultaneous prediction limit on the means of future samples from an exponential distribution // Technometrics. 1975. V. 17. P. 341–345.
186. *Nelson W. B.* Two-sample prediction, General Elect. Comp., TIS Rep., 68-c-40.
187. *Green J.* Asymptotic sample size for given confidence intervals length // JRSS. Sec. C. 1982. V. 31. P. 298–300.
188. *Нейман Ю.* Вводный курс теории вероятностей и математической статистики / Пер. с англ. — М.: Сов. Радио, 1962.
189. *Леман Э. Л.* Проверка статистических гипотез / Пер. с англ. — М.: Физматгиз, 1964.
190. *Чернов Г., Мозес Л. Е.* Элементарная теория статистических решений / Пер. с англ. — М.: Сов. Радио, 1962.
191. *Закс Ш.* Теория статистических выводов / Пер. с англ. — М.: Мир, 1975.
192. *Де Гроот М.* Оптимальные статистические решения. — М.: Мир, 1974.
193. *Pearson K.* Experimental discussion of the (χ^2, p) test for goodness-of-fit // Biometrika. 1932. V. 24. P. 351–381.
194. *Cochran W. G.* The chi-squared test of goodness-of-fit // AMS. 1952. V. 23. P. 315–345.
195. *Cochran W. G.* Some methods for strengthening the common chi-squared tests // Biometrika. 1954. V. 10. P. 417–451.
196. *Kallenberg W. C. M., Oosterhoff J., Schrever B. F.* The number of classes in chi-squared goodness-of-fit tests // JASA. 1985. V. 80, №392. P. 959–968.
197. *Mann H. B., Wald A.* On the choice of number of intervals in the application of the chi-square test // AMS. 1942. V. 18. P. 50–54.
198. *Щербинин А. Ф.* Об относительной эффективности критерия хи-квадрат и его аналогах // Надежность и контроль качества. 1986. №2. С. 13–17.
199. *Dahiya R. C., Gurland J.* How many classes in the Pearson chi-square test? // JASA. 1973. V. 68, №343. P. 707–712.
200. *Hamdan M. A.* The number and width of classes in the chi-square test // JASA. 1963. V. 58, №303. P. 678–679.
201. *Best D. J., Rayner J. C. M.* Are two classes enough for the goodness-of-fit test? // Statistica Neerlandica. 1981. V. 35. P. 157–163.
202. *Dahya R. C., Gurland J., Pearson K.* Chi-squares test of fit with random intervals // Biometrika. 1972. V. 59, №1. P. 147–153.
203. *Lawal H. B., Upton G. J. G.* An approximation to the distribution of the χ^2 goodness-of-fit statistic for the use with small expectations // Biometrika. 1980. V. 67, №2. P. 447–453.
204. *Larntz K.* Small-sample comparisons of exact levels for chi-square goodness-of-fit statistics // JASA. 1978. V. 73. P. 253–263.
205. *Yarnold J. K.* The minimum expectation in χ^2 goodness-of-fit tests and the accuracy of approximations for the accuracy of approximations for the null distribution // JASA. 1970. V. 65. P. 864–886.
206. *Wishart J.* χ^2 probabilities for large number of degrees of freedom // Biometrika. 1956. V. 43. P. 92–95.

207. *Романовский В. И.* Элементарный курс математической статистики. — М.-Л.: Госпланиздат, 1939.
208. *Barnett A., Eisen E.* A quartile test for differences in distribution // JASA. 1982. V. 77, № 377. P. 47–51.
209. *Орлов А. И.* Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат // Завод. лаб. 1985. Т. 51, № 1. С. 60–62.
210. *Kolmogorov A. N.* Confidence limits for an unknown distribution function // AMS. 1941. V. 12. P. 461–463.
211. *Смирнов Н. В.* Оценка расхождения между эмпирическими кривыми распределений в двух независимых выборках // Бюллетень МГУ. Сер. А. Вып. 2. 1939. С. 13–14.
212. *Stephens M. A.* Use of Kolmogorov–Smirnov, Cramer–von Mises and related statistics without extensive tables // JRSS. S. B. 1970. V. 32. P. 729–731.
213. *Chandra M., Singpurwalla N. D., Stephens M. A.* Kolmogorov statistics for tests of fit for the extreme-value and Weibull distributions // JASA. 1981. V. 76, № 375. P. 729–731.
214. *Koziol J. A., Byar D. P.* Percentage points of the asymptotic distributions of one and two sample K–S statistics for truncated or censored data // Technometrics. 1975. V. 17, № 4. P. 507–510.
215. *Stephens M. A.* EDF statistics for goodness-of-fit and some comparisons // JASA. 1974. V. 69. P. 730–737.
216. *Смирнов Н. В.* О распределении $n\omega^2$-критерия Мизеса // Математический сб. 1937. 2(44), № 5. С. 973–993.
217. *Смирнов Н. В.* О критерии Крамера–фон Мизеса // Успехи матем. наук (новая серия). 1949. Т. 4, № 4(32). С. 196–197.
218. *Мартынов Г. В.* Критерии омега-квадрат. — М.: Наука, 1978.
219. *Tiku M. L.* Chi-square approximation for the distributions of goodness-of-fit statistics U_N^2 and W_N^2 // Biometrika. 1965. V. 52, № 3–4. P. 630–633.
220. *Renyi A.* On the theory of order statistics // Acta Mathem. Acad. Scientiarum Hungarical. 1953. V. 4. P. 191–232.
221. *Залесский Б. А., Ольшевская О. В.* О функции распределения статистики омега-квадрат при малых выборках // Завод. лаб. 1989. № 7. С. 103–105.
222. *Goro Ishii.* On the exact probabilities of Renyi tests // Annal. Inst. Stat. Math. 1959. V. 2. P. 17–24.
223. *Anderson T. W., Darling D. A.* A test for goodness-of-fit // JASA. 1954. V. 49. P. 765–769.
224. *Lewis P. A. W.* Distribution of the Anderson–Darling statistic // AMS. 1961. V. 32. P. 1118–1123.
225. *Watson G. S.* Goodness-of-fit tests on a circle // Biometrika. 1961. V. 48, № 1–2. P. 109–114.
226. *Щербинин А. Ф.* Критерии согласия омега-квадрат по группированным наблюдениям // Надежность и контроль качества. 1983. № 1. С. 11–18.
227. *Sinclair C. D., Spurr B. D., Ahmad V. I.* Modified Anderson–Darling test // Commun. Stat.-Theor. Meth. 1990. V. 19, № 10. P. 3677–3686.
228. *Мардиа К.* Статистический анализ угловых наблюдений. — М.: Наука, 1978.
229. *Kuiper N. H.* Tests concerning random points on a circle // Proc. Konikl. Nederl. Akad. Van Wettenschappen. 1960. S. A. V. 63. P. 38–47.
230. *Stephens M. A.* The goodness-of-fit statistic V_n-distribution and significance points // Biometrika. 1965. V. 52. № 3–4. P. 309–321.
231. *O'Reilly F. J., Stephens M. A.* Characterization and goodness-of-fit tests // JRSS. 1982.
232. *Darling J.* The Kolmogorov–Smirnov, Cramer–von Mises tests // AMS. 1957. V. 28. P. 823–838.
233. *Durbin J.* Some methods of constructing exact tests // Biometrika. 1961. V. 48, № 1–2. P. 41–57.
234. *Pearson K.* On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drown of random // Biometrika. 1933. V. 25. P. 379–410.

235. *Pearson E. S.* The probability integral transformation for testing goodness-of-fit and combining independent tests of significance // Biometrika. 1939. V. 30. P. 134–148.
236. *Katzenbeisser W., Hackl P.* An alternation to the Kolmogorov–Smirnov two-sample test // Commun. Stat.-Theor. Meth. 1986. V. 15, №4. P. 1163–1177.
237. *Anderson T. W.* On the distribution of the two-sample Cramer–von Mises criterion // AMS. 1962. V. 33. P. 1148–1159.
238. *Frozini B. V.* A survey of a class of goodness-of-fit statistics // Metron. 1978. V. 36, №1–2. P. 3–49.
239. *Frozini B. V.* On the distribution and power of a goodness-of-fit statistic with parametric and nonparametric applications, „Goodness-of-fit" / Ed. by Revesz P., Sarkadi K., Sen P. K., Amsterdam–Oxford–New York: North–Holland. Publ. Comp., 1987, P. 133–154.
240. *Shapiro S. S., Wilk M. B.* An analysis of variance test for normality (complete samples) // Biometrika. 1965. V. 52, №3. P. 591–611.
241. *Lloyd E. N.* Least-squares estimation of location and scale parameters using order statistics // Biometrika. 1952. V. 39. P. 88–95.
242. *Shapiro S. S., Wilk M. B.* Approximations for the null distribution of the W statistic // Technometrics. 1968. V. 10, №4. P. 861–866.
243. *Shapiro S. S., Wilk M. B., Chen H. J.* A comparative study of various tests for normality // JASA. 1968. V. 63, №324. P. 1343–1372.
244. *Shapiro S. S., Francia R. S.* An approximate analysis of variance test normality // JASA. 1972. V. 67, №337. P. 215–216.
245. *Weisberg S., Bingham C.* An approximate analysis of variance test for non-normality suitable for machine calculation // Technometrics. 1975. V. 17, №1. P. 133–134.
246. *Royston J. P.* Correcting the Shapiro–Wilk test W for ties // J. Stat. Comput. Simul. 1989. V. 31, №4. P. 237–249.
247. *Vasicek O.* A test for normality based on sample entropy // JRSS. 1976. V. 38, №1. P. 54–59.
248. *Prescott P.* On a test for normality based on sample entropy // JRSS. 1976. V. 38, №3. P. 254–256.
249. *Hegazy Y. A. S., Green J. R.* Some new goodness-of-fit tests using order statistics // Appl. Statist. 1975. V. 24, №3. P. 299–308.
250. *Csörgő H., Revesz P.* Quantile processes and sums of weighted spasing for composite goodness-of-fit. In: Statistics and Related Topics, North Holland, Amsterdam, N. Y., 1981. P. 69–87.
251. *Aly E.-E., Csörgő M.* Quadratic nuisance parameter-free goodness-of-fit tests in the presence of location and scale parameters // CJS. 1985. V. 13. P. 53–70.
252. *Aly E.-E., Shayib M. A.* On some goodness-of-fit tests for the normal, logistic and extreme-value distributions // Commun. Stat.-Theor. Meth. 1992. V. 21, №5. P. 1297–1308.
253. *Filliben J. J.* The probability plot correlation coefficient test for normality // Technometrics. 1975. V. 17, №1. P. 111–117.
254. *Gulledge T. R., Looney S. W.* An alternative test for normality Comput. Sci. And Statist // Proc. 16-th Symp. Interface, Atlanta, Ga, Marth 1984, Amsterdam e.a., 1985. P. 265–267.
255. *La Brecque J.* Goodness-of-fit tests based on nonlinearity in probability plots // Technometrics. 1977. V. 19, №3. P. 293–306.
256. *Locke C., Spurrier J. D.* The use of U-statistics for testing normality against nonsymmetric alternative // Biometrika. 1976. V. 63, №1. P. 143–147.
257. *Locke C., Spurrier J. D.* The use of U-statistics for testing normality against alternatives with both tails heavy or both tails light // Biometrika. 1977. V. 64, №3. P. 638–640.
258. *Oja H.* Two location and scale goodness-of-fit tests // Biometrika. 1981. V. 68, №3. P. 637–640.
259. *Davids C. E., Quade D.* U-statistics for skewness or symmetry // Commun. Stat.-Theor. Meth. 1978. V. A7, №5. P. 413–418.
260. *Oja H.* New tests for normality // Biometrika. 1983. V. 70, №1. P. 279–299.
261. *Geary R. C.* The ratio of the mean deviation to the standard deviation as a test of normality // Biometrika. 1935. V. 27. P. 310–322.

262. *Geary R. C.* Moments of the ratio of the mean deviation to the standard deviation for normal samples // Biometrika. 1936. V. 28. P. 295–307.
263. *Geary R. C.* Testing for normality // Biometrika. 1947. V. 34. P. 209–242.
264. *David H. A.*, *Hartley H. O.*, *Pearson E. S.* The distribution of the ratio, in a single normal sample, of range to standard deviation // Biometrika. 1954. V. 41. P. 482–493.
265. *Pearson E. S.*, *Stephens M. A.* The ratio of range to standard deviation in the same normal sample // Biometrika. 1964. V. 512, № 3–4. P. 484–487.
266. *Thomson G. W.* Bounds for the ratio of range to standard deviation // Biometrika. 1985. V. 42, № 1–2. P. 268–269.
267. *Spiegelhalter D. J.* A test for normality against symmetric alternatives // Biometrika. 1977. V. 64, № 2. P. 415–418.
268. *Spiegelhalter D. J.* An omnibus test for normality for small samples // Biometrika. 1980. V. 67, № 2. P. 493–496.
269. *Sarkadi K.* On testing for normality, A Magyar Tud. Akad., Matem. Kutato Intezet. Kozlemenyei., V. A5, 1960. P. 269–275.
270. *Золотухина Л. А.*, *Винник Е. В.* Эмпирическое исследование мощности критерия Саркади и его модификации // Завод. лаб. 1985. Т. 52, № 1. С. 51–55.
271. *Kawata T.*, *Sakamoto H.* On the characterization of the normal population by the independence of the sample mean and the sample variance // J. Math. Soc. Japan. 1949. № 1. P. 111–115.
272. *Заингер А. А.* О независимых выборках из нормальной совокупности // Успехи матем. наук. 1951. Т. VI. Вып. 5. С. 172–175.
273. *Коган А. М.*, *Ландеман Э. М.* Информационный аналог теоремы о независимости среднего и дисперсии // Проблемы устойчивости стохастических систем: Тр. семинара. — М.: ВНИИСИ, 1985. — С. 763–767.
274. *Lin Ch.-Ch.*, *Mudholkar G. S.* A simple test for normality against asymmetric alternatives // Biometrika. 1980. V. 67, № 2. P. 455–461.
275. *Nelson B. B.* Testing for normality // J. of Quality Technology. 1983. V. 15, № 3. P. 141–143.
276. *Nelson L. S.* A simple test for normality // J. of Quality Technology. 1983. V. 13. P. 76–77.
277. *Mudholkar G. S.*, *Lin C. C.* On two applications characterization theorems to goodness-of-fit // Goodness-of-fit, Amsterdam, 1987. P. 395–414.
278. *Martinez J.*, *Iglewicz B.* A test for departure from normality based on a biweight estimator of scales // Biometrika. 1981. V. 68, № 1. P. 331–333.
279. *D'Agostino R. B.* An omnibus test of normality for moderate and large size samples // Biometrika. 1971. V. 58, № 2. P. 341–348.
280. *D'Agostino R. B.* Small sample probability points for the D-test of normality // Biometrika. 1972. V. 59, № 1. P. 219–221.
281. *Dawton F.* Linear estimates with polynomial coefficients // Biometrika. 1966. V. 53. P. 129–141.
282. *D'Agostino R. B.* Monte Carlo power comparisons of the W' and D tests of normality // Common. Stat. 1973. V. 1. P. 545–551.
283. *Pearson E. S.* A further development of tests for normality // Biometrika. 1930. V. 22. P. 239–249.
284. *D'Agostino R. B.*, *Pearson E. S.* A further development of tests departure from normality. Empirical results for the distribution of b_2 and $\sqrt{b_1}$ // Biometrika. 1973. V. 60, № 3. P. 613–622.
285. *D'Agostino R. B.* Transformation to normality of the null distribution of g_1 // Biometrika. 1970. V. 57. P. 679–681.
286. *Anscombe F. J.*, *Glynn W. J.* Distribution of the kurtosis statistic b_2 for normal samples // Biometrika. 1983. V. 70, № 1. P. 227–234.
287. *Bowman K. O.*, *Stenton L. R.* Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 // Biometrika. 1975. V. 62, № 2. P. 243–249.
288. *Littele R. C.*, *Folks J. L.* Asymptotic optimality on Fisher's method of combining independent tests. 11 // JASA. 1973. V. 68. P. 193–194.

289. *Jarque C. M., Bera A. K.* A test for normality of observation and regression residuals // Internat. Stat. Review. 1987. V. 55, № 2. P. 163–172.
290. *Fenerverger A., Mureika R. A.* The empirical characteristic function and its application // Annal. Statist. 1977. V. 5. P. 88–97.
291. *Heathote C. R.* A test of goodness-of-fit for symmetric random variables // Austr. J. Statist. 1972. V. 14. P. 172–181.
292. *Kontrouvelis I. A.* A goodness-of-fit test of simple hypothesis based on the empirical characteristic function // Biometrika. 1980. V. 67. P. 138–240.
293. *Murota K., Takeuchi K.* The Studentized empirical characteristic function and its application to test for the shape of distribution // Biometrika. 1981. V. 68. P. 55–65.
294. *Wilk M. B., Shapiro S. S.* The joint assessment of normality of several independent samples // Technometrics. 1968. V. 10, № 10. P. 825–839.
295. *Винник Е. В., Золотухина Л. А.* Применение модифицированного критерия Саркади для проверки гипотезы нормальности по совокупности малых выборок // Завод. лаб. 1987. Т. 53, № 7. С. 51–54.
296. *Wald A., Wolfowitz J.* An exact test of randomness in the nonparametric case based on serial correlation // AMS. 1943. V. 14. P. 378–388.
297. *Арлей Н., Бух К.* Введение в теорию вероятностей и математическую статистику / Пер. с англ. — М.: ИЛ, 1951.
298. *Shapiro S. S., Wilk M. B.* An analysis of variance test for the exponential distribution (complete samples) // Technometrics. 1972. V. 14. P. 355–370.
299. *Samanta M., Schwarz C. J.* The Shapiro–Wilk test for exponentiality based on censored data // JASA. 1988. V. 83, № 402. P. 528–531.
300. *Stephens M. A.* On the W test for exponentiality with origin known // Technometrics. 1978. V. 20. P. 353–357.
301. *Spinelli J. J., Stephens M. A.* Tests for exponentiality when origin and scale parameters are unknown // Technometrics. 1987. V. 29, № 4. P. 471–476.
302. *Spurrier J. D.* On overview of tests of exponentiality // Commun. Stat.-Theor. Meth. 1984. V. 13. P. 1635–1654.
303. *Sarhan A. E.* Estimation of the mean and standard deviation by order statistics // 1954. AMS. V. 25. P. 317–328.
304. *Pettit A. N.* Tests for the exponentiality distribution with censored data using Cramer–von Mises statistics // Biometrika. 1977. V. 64, № 3. P. 629–632.
305. *Brain C. W., Shapiro S. S.* A regression test for exponentiality: censored and complete samples // Technometrics. 1983. V. 25, № 1. P. 69–76.
306. *Kimber A. C.* Tests for the exponential, Weibull and Gumbel distributions based on the stabilized probability plot // Biometrika. 1985. V. 72, № 3. P. 661–663.
307. *Michael J. R.* The stabilized probability plot // Biometrika. 1983. V. 70. P. 11–17.
308. *Coles S. G.* On goodness-of-fit tests for the two-parameter Weibull distribution derived from the stabilized probability plot // Biometrika. 1989. V. 76, № 3. P. 593–598.
309. *Moran P. A. P.* The random division of an interval, 11 // JRSS. 1951. V. 13. P. 147–150.
310. *Кокс С., Льюис П.* Статистический анализ последовательностей событий / Пер. с англ. — М.: Мир, 1969.
311. *Klimko L. A., Antle C. E., Rademaker A. W., Rockette H. E.* Upper bounds for the power of invariant tests for the exponential distribution with Weibull alternative // Technometrics. 1975. V. 17, № 3. P. 357–360.
312. *Tanaka S., Ichikawa M.* Approximate formula of coefficient of variation for Weibull distribution // Reliability Engineering. 1983. V. 4, № 3. P. 141–143.
313. *Hollander M., Proshan F.* Testing whether new is better than used // AMS. 1972. V. 43. P. 1136–1146.
314. *Sturges H. A.* The choice of a class interval // JASA. 1926. V. 21. P. 65–66.
315. *Kochar S. C.* Testing exponentiality against monotone failure rate average // Commun. Stat.-Theor. Meth. 1985. № 2. P. 381–392.
316. *Klefsjö B.* Some tests against aging based on the total time on test transform // Commun. Stat.-Theor. Meth. 1983. V. 14, № 12. P. 917–927.

317. *Deshande J. V.* A class on tests for exponentiality against increasing failure rate average alternative // Biomertika. 1983. V. 70. P. 514–518.
318. *Barlow R. E.* Likelihood ratio tests for restricted families of probability distributions // AMS. 1968. V. 39. P. 547–560.
319. *Bickel P. J.*, *Doksum K.* Tests for monotone failure rate based on normalized spacing // 1969. AMS. V. 40. P. 1216–1235.
320. *Klefsjö B.* Testing exponentiality against HNBUE // Scand. J. Statist. 1983. V. 10. P. 67–75.
321. *Hollander M.*, *Proshan F.* Tests for the mean residual life. A methods and corrections // Biometrika. 1980. V. 67. P. 259–261.
322. *Epps T. W.*, *Pulley L. B.* A test of exponentiality vs. monotone–hazard alternatives from the empirical characteristics function // JRSS. Sec. B. 1986. V. 48, № 2. P. 206–216.
323. *Bergman B.* On the replacement and the total time on test concept // Scand. J. Statist. 1979. V. 6. P. 161–168.
324. *Bergman B.* Crossings in the total time on test plot // Scand. J. Statist. 1977. V. 4. P. 171–177.
325. *Barlow R. E.*, *Campo R.* Total time on test processes and applications to failure data analysis // Reliability and fault free analysis (ed. Barlow, Fussel and Singpurwalla), SIAM, Philadelphia, 1975.
326. *Dipankar Bandyopadhyay*, *Basu A. P.* A note on tests for exponentiality by Deshpande // Biometrika. 1989. V. 76, № 2. P. 403–405.
327. *Sherman B.* A random variable related to the spacing of sample values // AMS. 1950. V. 21, № 3. P. 339–361.
328. *Sherman B.* Percentiles of the ω_n statistic // AMS. 1957. V. 28, № 1. P. 257–261.
329. *Bartholomew D. J.* Note on the use of Sherman's statistics as a test for randomness // Biometrika. 1954. V. 41. P. 556–558.
330. *Hartley H. O.* The maximum F-ratio as a short-cut test of heterogeneity of variance // Biometrika. 1950. V. 37. P. 308–312.
331. *Greenwood V.* The statistical study of infection disease // JRSS. Sec. A. 1946. V. 109. P. 85–110.
332. *Stephens M. A.* Further percentage points for Greenwood's statistics // JRSS. Sec. A. 1981. V. 144. P. 364–366.
333. *Burrows P. M.* Selected percentage points of Greenwood's statistics // JRSS. Sec. A. 1981. V. 142. P. 256–258.
334. *Deshpande J. V.* A class of tests for exponentiality against increasing failure rate average alternatives // Biometrika. 1983 V. 70. P. 514–518.
335. *Engelhardt M. E.*, *Bain L. J.* Uniformly more powerful unbiased tests for the parameters of the gamma distribution // Theory and Applications of Reliability. V. 1. N. Y.: Acad. Press. — P. 307–314.
336. *Lawless J. F.* Statistical models and methods for lifetime data. — N. Y.: J. Welley, 1982.
337. *Keating J. P.*, *Glaser R. E.*, *Ketchum N. S.* Testing hypotesis about the shape parameter of a gamma distribution // Technometrics. 1990. V. 32, № 1. P. 67–82.
338. *Kimball B. F.* Some basic theorems for developing tests of fit for the case of the non-parametric probability distribution function. 1 // AMS. 1947. V. 18. P. 540–548.
339. *Moran P. A. P.* The random division of an intervals // JRSS. 1947. Sec. B. V. 9. P. 92–98.
340. *Smirnoff N.* Sur la distribution de $n\omega^2$ // Compte Rendus de l'Academie des Sciences. Paris, 1932. № 202. P. 449.
341. *Cheng R. C. H.*, *Thornton K. M.* Selected percentage points of the Moran statistic // J. Stat. Comput. Simul. 1988. V. 30. P. 189–194.
342. *Cheng R. C. H.*, *Stephens M. A.* A goodness-of-fit using Moran's statistic with estimated parameters // Biometrika. 1989. V. 76, № 2. P. 385–392.
343. *Cheng S. W.*, *Spiring F. A.* A test to identify the uniform distribution with applications to probability plotting and other distributions // IEEE Trans. Reliability. 1987. V. R–36, № 1. P. 98–105.

344. *Kosik P., Sarkadi K.* A new goodness-of-fit test // Proc. of 5-th Pannonian Symp. of Math. Stat., Visegrad, Hungary, 20–24 May, 1985. P. 267–272.
345. *Dudewicz E. J., van der Meulen E. C.* Entropy-based tests of uniformity // JASA. 1981. V. 76, № 376. P. 967–974.
346. *Neyman J.* „Smooth“ tests for goodness-of-fit // Scand. Aktuarietidsrift. 1937. V. 20. P. 149–199.
347. *Young D. L.* The linear nearest neighbour statistic // Biometrika. 1982. V. 69, № 2. P. 477–480.
348. *Clark P. J., Evans F. C.* Distance to nearest neighbour as a measure of spatial relationships in populations // Ecology. 1954. V. 35. P. 445–457.
349. *Pinder D. A., Witherick M. E.* Nearest-neighbour analysis of linear point patterns // Tijdshrift voor Economischl th social Geografie. 1973. V. 64. P. 160–163.
350. *Quesenberry C. P., Miller F. L.* Power studies of some tests for uniformity // J. Statist. Comput. Simul. 1977. V. 5. P. 169–191.
351. *David F. N.* On Neyman's „smooth“ test for goodness-of-fit // Biometrika. 1939. V. 31. P. 191–199.
352. *Miller F. L., Quesenberry C. P.* Power studies of some tests for uniformity, 11 // Commun. Stat.–Simul. Comput. 1979. S. B. V. 8, № 3. P. 271–290.
353. *Solomon H., Stephens M. A.* On Neyman's statistics for testing uniformity // Commun. Stat.-Simul. Comput. 1983. V. 12. P. 127–134.
354. *Смирнов Н. В.* О критерии симметрии закона распределения случайной величины // ДАН СССР. 1947. Т. 56, № 1. С. 13–16.
355. *Wilcoxon F.* Individual comparisons by ranking methods // Biometrics, Bull. 1945. V. 1. P. 80–83.
356. *Antille A., Kersting G., Zucchini W.* Testing symmetry // JASA. 1982. V. 77, № 379. P. 639–646.
357. *Bhattacharya P. K., Gastwirth J. L., Wright A. L.* Two modified Wilcoxon tests for symmetry about an unknown location parameters // Biometrika. 1982. V. 69, № 2. P. 377–382.
358. *Antille A., Kersting G.* Tests for symmetry // Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete. 1977. V. 39. P. 235–255.
359. *Finch S. J.* Robust univariate test of symmetry // JASA. 1977. V. 72, № 358. P. 387–392.
360. *Boos D. D.* A test for symmetry associated with the Hodges–Lehmann estimator // JASA. 1982. V. 77, № 379. P. 647–651.
361. *Gupta M. K.* An asymptotically nonparametric test of symmetry // AMS. 1967. V. 38. P. 849–866.
362. *Moses L. E.* Query: Confidence limits from rank tests // Technometrics. 1965. V. 7. P. 257–260.
363. *Frazer D. A. S.* Most powerfull rank-type tests // AMS. 1957. V. 28. P. 1040–1043.
364. *Klotz J.* Smoll sample power and efficiency for the one sample Wicoxon and normal scores test // AMS. 1963. V. 34. P. 624–632.
365. *Гаек Я., Шидак З.* Теория ранговых критериев / Пер. с англ. — М.: Наука, 1971.
366. *Johnson N. L.* Systems of frequency curves generated by methods of translation // Biometrika. 1946. V. 36. P. 146–148.
367. *Висленев Ю. С., Самосейко В. Ф.* О выборе закона распределения по моментам случайной величины // Надежность и контроль качества. 1987, № 5. С. 21–24.
368. *Бостанджиян В. А.* Распределение Джонсона (препринт), ИХФ АН СССР, Черноголовка, 1978.
369. *Johnson N. L.* Tables to facilitate fitting S_U frequency curves // Biometrika. 1965. V. 52. P. 547–558.
370. *Демаков И. П., Потепун В. Е.* Графо-аналитический метод построения полуэмпирических функций распределения малых выборок // Труды метрологических институтов СССР. Общие вопросы метрологии: Сб. — Л.: ВНИИМ, 1972. — С. 96–102.
371. Статистические методы оценки качества и надежности промышленных изделий по результатам малого числа испытаний: Методологические указания. — Уфа: Изд-во УАИ им. С. Орджоникидзе, 1983.

372. *Демаков И. П.* Проблема принятия решения по малому числу наблюдений (обзор состояния и перспективы) // Методы статистического анализа и обработки малого числа наблюдений при контроле качества и надежности приборов и машин: Сб. — Л.: ЛДНТП, 1974. — С. 4–18.
373. *Гаскаров Д. В.*, *Шаповалов В. И.* Малая выборка. — М.: Статистика, 1978.
374. *Королькова Л. И.* Непараметрическое оценивание функции распределения по малой выборке. — Челябинск, 1988, ВИНИТИ, № 6602–888.
375. *Корольков И. В.*, *Королькова Л. И.* Оценивание по малой выборке с использованием бета-распределения // Проблемы разработки и использования гибких автоматизированных производств на предприятиях уральского региона: Сб. — Свердловск: УНЦ АН СССР. — С. 12.
376. *Горский Л. К.* Статистические алгоритмы исследования надежности. — М.: Наука, 1970.
377. Алгоритмы и программы восстановления зависимостей / Под ред. В. Н. Вапника. — М.: Наука, 1984.
378. *Еременко И. В.*, *Свердлик А. Н.* Об одном методе построения законов распределения величин при малом числе испытаний // Некоторые вопросы специального применения вычислительной техники: Сб. — Л.: ЛВИКА им. А. Ф. Можайского, 1963. — С. 18–29.
379. *Демидович Б. П.*, *Марон И. А.*, *Шувалова Э. З.* Численные методы анализа. — М.: Физматгиз, 1963.
380. *Попов С. А.*, *Шаповалов В. И.* Об оценивании закона распределения по ограниченным данным // Надежность и контроль качества. 1982. № 11. С. 29–32.
381. *Дружинин Г. В.*, *Воронова О. В.* Сравнение методов построения эмпирической функции распределения по малому числу наблюдений // Надежность и контроль качества. 1983. № 1. С. 11–18.
382. *Чавчанидзе В. В.*, *Кумсенишвили В. А.* Об определении законов распределения на основе малого числа наблюдений // Применение вычислительной техники для автоматизации производства. — М.: Машгиз, 1961.
383. *Гаскаров Д. В.*, *Голинкевич Т. А.*, *Мозгалевский А. В.* Прогнозирование технического состояния и надежности радиоэлектронной аппаратуры. — М.: Сов. радио, 1974.
384. *Гаскаров Д. В.*, *Мозгалевский А. В.*, *Шаповалов В. И.* Некоторые вопросы прогнозирования по ограниченной информации // Электронная техника. 1975, с. 8, вып. 4(34). С. 15–20.
385. *Диттл Р. Дж. А.*, *Рубин Д. Б.* Статистический анализ данных с пропусками / Пер. с англ. — М.: Финансы и статистика, 1991.
386. *Cochran W. G.*, *Cox C. M.* Experimental design. N. Y.: J. Willey a. Sons, 1957.
387. *Satterwaite R.* An approximate distribution of estimates of variance components // Biometrics. 1946. V. 2. P. 110.
388. *Welch B. L.* The generation of „Student's“ problems when several different population variance are involved // Biometrika. 1947. V. 34. P. 28–35.
389. *Lord E.* The use of range in place of standard deviation, in the t-test // Biometrika. 1947. V. 34. P. 41–67.
390. *Daly J. F.* The use of the sample range in an analogue of Student's t-test // AMS. 1946. V. 17. P. 71–74.
391. *Walsh J. E.* Some significance tests based on order statistics // AMS. 1946. V. 17, № 1. P. 44–52.
392. *Wolfe D. A.* Two-stage two-sample median test // Technometrics. 1977. V. 19, № 4. P. 495–501.
393. *Артемьев Е. Ю.*, *Мартынов Е. М.* Вероятностные методы в психологии. — М.: Изд-во МГУ, 1975.
394. *Paulson E.* On optimum solution to the k-sample slippage problem for the normal distribution // AMS. 1952. V. 23, 4. P. 610–616.
395. *Baily R.* Tables of the Bonferroni t statistic // JASA. 1977. V. 72, № 358. P. 469–478.

396. *Nelson P. R.* Exact critical points for the analysis of means // Commun. Stat.-Theor. Meth. 1982. V. 11, №6. P. 699–709.
397. *Hochberg Y.* Some generalization of the T–method in simultaneous inference // J. Multivar. Anal. 1974. №4. P. 224–234.
398. *Stoline M. R.*, *Ury H. K.* Tables of the studentized maximum modules distributions and an application to multiple comparisons amond means // Technometrics. 1978. V. 21, №1. P. 87–93.
399. *Half G. J.*, *Hendricson R. W.* A table of percentage points of the largest absolute value of k Student t variates and its applications // Biometrika. 1971. V. 58. P. 323–332.
400. *Pillai K. C. S.*, *Ramachandran K. V.* On the distribution of the ratio of the i-th observation in an ordered sample from a normal population to an independent estimate of the standard deviation // AMS. 1954. №3. P. 565–572.
401. *Шеффе Г.* Дисперсионный анализ. — М.: Физматгиз, 1963.
402. *Newman D.* The distribution of the range in samples from normal populations, expressed in terms of an independent estimate of standard deviation // Biometrika. 1939. V. 31. P. 20–30.
403. *Keuls M.* The use of the studentized range in connection with an analysis of varianse // Euphytica. 1952. V. 1. P. 112–122.
404. *Duncan D. B.* Multiple range and multiple F test // Biometrics. 1955. V. 11. P. 1–42.
405. *Link R. F.*, *Wallace D. L.* Some short cuts to allowances, Princeton Univ., March, 1957.
406. *Kurtz T. E.*, *Link R. F.*, *Tukey J. W.*, *Wallace D. L.* Short-cut multiple comparisons for balanced single and double classification: part 1, Results // Technometrics. 1965. V. 7. P. 95–161.
407. *Pearson E. S.* The analysis of variance in classes of non-normal variation // Biometrika. 1931. V. 23. P. 114–134.
408. *Geary R. C.* Testing for normality // Biometrika. 1947. V. 23. P. 209–241.
409. *Gayen A. K.* The distribution of the variance ratio in random samples of any size drawn from non-normal universes // Biometrika. 1950. V. 37. P. 236–255.
410. *Arizono I.*, *Ohta M.* A test of homogeneity of variances based on sample entropy // Bull. Osaka Prefect. A, v. 36, №1, 1987. P. 29–37.
411. *Bartlett M. S.* Properties of sufficiency of statistical tests // Proc. Roy. Soc. 1937. A 160. P. 268–287.
412. *Box G. E. P.* A general distribution theory for a class of likelihood criteria // Biometrika. 1949. V. 36. P. 317–346.
413. *Wallis W. A.*, *Moore G. H.* A significance test for time series analysis // JASA. 1941. V. 36. P. 401–409.
414. *Cochran W. G.* The distribution of the largest of a set of estimated variances as a fraction of their total // Annals of Eugenics. 1941. V. 11. P. 47–52.
415. *Bliss C. I.*, *Cohran W. G.*, *Tukey J. W.* A rejection criterion based upon the range // Biometrika. 1956. V. 43. P. 418–422.
416. *Hartley H. O.* The maximum F-ratio as a short-cut test of heterogeneity of variance // Biometrika. 1950. V. 37. P. 308–312.
417. *Cadwell J. H.* Approximating to the distributions of measures of dispersion by a power of χ^2 // Biometrika. 1952. V. 40. P. 336–346.
418. *Leslie R. T.*, *Brown B. M.* Use of range in testing heterogeneity of variance // Biometrika. 1966. V. 53. P. 221–227.
419. *Samiuddin M.*, *Atiqullah M.* A test for equality of variance // Biometrika. 1976. V. 63, №1. P. 206–208.
420. *Samiuddin M.*, *Hanif M.*, *Asad H.* Some comparisons of the Bartlett and cube root tests of homogeneity of variance // Biometrika. 1978. V. 65, №1. P. 218–221.
421. *David H. A.* The ranking of variance in normal population // JASA. 1956. V. 51. P. 112–116.
422. *Кэндэлл М.* Ранговые корреляции. — М.: Статистика, 1975.
423. *Nagarsenker P. B.* On a test of equality of several exponential survival distributions // Biometrika. 1980. V. 67, №2. P. 475–478.

424. *Chen H. J.* A new range statistic for comparisons of several exponential location parameters // Biometrika. 1982. V. 69, №1. P. 257–260.
425. *Sukhatme P. V.* On the analysis of k sample from exponential populations with especial reference to the problems of random intervals // Statist. Reseach. Memoric. 1936. V. 1. P. 94–112.
426. *Epstein B., Tsao C. K.* On the tests based on ordered observation from two exponential populations // AMS. 1953. V. 24. P. 458–466.
427. *Perng S. K.* A test for equality of two exponential distributions // Statist. Neerlandica. 1978. V. 32. P. 93–102.
428. *Hseih H. K.* On testing the equality of two exponential distributions // Technometrics. 1981. V. 23. P. 265–269.
429. *Hogg R. V., Tanis E. A.* An iterated procedure for testing the equality of several exponential distributions // JASA. 1963. V. 58. P. 435–443.
430. *Singh N., Nagaran P.* The likelihood ratio test for the equality of k [$\geqslant 2$] two parameters exponential distributions based of type 2 censored samples // J. Stat. Comput. Simul. 1983. V. 18. P. 373–381.
431. *Epstein B., Sobel M.* Some theorems relevant to life testing from on exponential populations // AMS. 1954. V. 25. P. 373–381.
432. *Singh N.* A simple and asymptotically optimal test for the equality of k [$\geqslant 2$] exponential distributions based on type 2 censored samples // Commun. Stat.-Theor. Meth. 1985. V. 14, №7. P. 1615–1625.
433. *Вальд А.* Последовательный анализ / Пер. с англ. — М.: Физматгиз, 1960.
434. *Башаринов А. Е., Флейшман Б. С.* Об эффективности метода последовательного анализа в устройствах обнаружения слабых сигналов в шумах // Радиотехн. и электрон. 1958. Т. 3, №8. С. 42–47.
435. *Крапивин В. Ф.* Таблицы распределения Вальда. — М.: Наука, 1965.
436. *Mann H. B., Whitney D. R.* On a test of whether one of two random variables is stochastically larger than the other // AMS. 1947. V. 18. P. 50–60.
437. *Singh N.* The ratio test for the parameters exponential distributions // Commun. Stat.-Theor. Meth. 1985. V. 13, №6. P. 116–119.
438. *Iman R. L.* An approximation to the exact distribution of the Wilcoxon–Mann–Whitney rank sum test statistic // Commun. Statist. 1976. A5. P. 587–598.
439. *Iman R. L.* Use of a t-statistic as an approximation to the exact distribution of the Wilcoxon signed rank test statistic // Commun. Statist. 1974. V. 3. P. 795–806.
440. *Fisher R. A., Yates F.* Statistical tables for biological, agricultural and medical research. Edinburgh & London: Oliver and Boyd, 1946.
441. *Terry M. E.* Some rank order test which are most powerfull against specific parametric alternatives // AMS. 1952. V. 23. P. 346–366.
442. *Hoeffding W.* Optimum non-parametric tests // Proc. 11-th Berkley Symp., 1950. P. 83–92.
443. *Šidak Z.* Tables for two normal-scores rank tests for two-sample location problem // Appl. Math. 1973. V. 18, №5. P. 333–345.
444. *Бернстейн Ф.* Сборник статистических таблиц / Пер. с англ. — М.: Статистика, 1968.
445. *Mosteller F.* A k-sample test for an extreme population // AMS. 1948. V. 19. P. 58–65.
446. *Rosenbaum S.* Tables for a nonparametric test of location // AMS. 1954. V. 25. P. 146–150.
447. *Haga T.* A two-sample rank test on location // Annal. Inst. Stat. Math. 1959/60. V. 11. P. 211–219.
448. *Hojek S.* Tables for the two-sample Haga test of location // Aplikate Mat. 1978. V. 23, №4. P. 237–247.
449. *Šidak Z., Vondrauek J.* A sample non-parametric test of the difference in location of two populations // Appl. Mat. 1957. V. 2. P. 215–221.
450. *Šidak Z.* Tables for the two-sample location E-test based on exceeding observations // Appl. Mat. 1977. V. 22, №3. P. 166–175.
451. *Kruskal W. H., Wallis A.* Use of ranks in one criterion variance analysis // JASA. 1952. V. 47. P. 583–621.

452. *Iman R. L., Davenport J. M.* New approximations to the exact distribution at the Kruskal–Wallis test statistic // Commun. Stat.-Theor. Meth. 1976. V. 5. P. 1335–1348.
453. *Nemenyi P.* Distribution-free multiple comparisons. New York, State Univ., Downstate Medical Center, 1963.
454. *Wilcoxon F., Wilcox R. A.* Some rapid approximate statistical procedures, Lederle Laboratories, Ptarl River, New York, 1964.
455. *Хеттманспергер Т.* Статистические выводы, основанные на рангах / Пер. с англ. — М.: Финансы и статистика, 1987.
456. *Terpstra T. J.* The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking // Indagationes Math. 1952. V. 14. P. 327–333.
457. *Jonckheere A. R.* A distribution-free k-sample test against ordered alternatives // Biometrika. 1954. V. 41. P. 133–145.
458. *Odeh R. E.* On Jonckheere's k-sample test against ordered alternatives // Technometrics. 1971. V. 13, № 4. P. 912–918.
459. *Lewis J. L.* A k sample test based on range intervals // Biometrika. 1972. V. 59, № 1. P. 155–160.
460. *Bhapkar V. P., Deshpande J. V.* Some nonparametric tests for multisample problems // Technometrics. 1968. V. 10, № 3. P. 578–585.
461. *Deshpande J. V.* Some nonparametric tests of statistical hypotheses, Dissertation submitted for the Ph. D. degree to the Univ. of Poona, 1965.
462. *Deshpande J. V.* A nonparametric test based on U-statistics for the problem of several samples // J. Indian Statist. Assoc. 1965. V. 3. P. 20–29.
463. *Barbur A. D., Cartwright D. I., Donnely J. B., Eagleson G. K.* A new rank test for the k-sample problem // Commun. Stat.-Theor. Meth. 1985. V. 14, № 6. P. 1471–1484.
464. *Usawa H.* Locally most powerfull rank tests for two-sample problems // AMS. 1960. V. 31. P. 685–702.
465. *Crouse C. F.* Distribution free test based on the sample distribution function // Biometrika. 1966. V. 53. P. 99–108.
466. *Page E. B.* Ordered hypotheses for multiple treatments: A significance test for linear ranks // JASA. 1963. V. 58. P. 216–230.
467. *Frideman M.* The use of ranks to avoid the assumption of normality implicit in the analysis of variance // JASA. 1937. V. 32. P. 675–701.
468. *Kendall M. G., Babington Smith.* The problem of m rankings // AMS. 1939. V. 10. P. 275–287.
469. *Iman R. L., Davenport J. M.* Approximations of the critical region of the Friedman statistic // Commun. Stat.-Theor. Meth. 1980. V. 9. P. 571–595.
470. *Anderson R. L.* Use of contingency tables in the analysis of consumer preference // Biometrics. 1959. V. 15. P. 582–590.
471. *Kanneman K.* An incidence test for k related samples // Biom. Zeitschrift, ed. 18, 1976. S. 3–11.
472. *Schach S.* An alternative to the Friedman test with certain optimality proporties // Annal. of Statistics. 1979. V. 7, P. 537–550.
473. *Quade D.* Analyzing randomized blocks by weighted rankings, Report SW 18/72, Math. Center Amsterdam, 1972.
474. *Quade D.* Using weighted ranking in the analysis of complete blocks with additive block effects // JASA. 1979. V. 74, № 370. P. 680–683.
475. *Quade D., Silva C.* Evaluting of weighted rankings using expected significance level // Commun Stat.-Theor. Meth. 1980. V. 9. P. 1087–1096.
476. *Ehrenberg A. S. C.* On sampling from a population of a rankers // Biometrika. 1952. V. 39. P. 82–87.
477. *Alvo M., Cabilio P., Feigin P. D.* Asimptotic theory for measures of concordance with special reference to average Kendall tau // Annal. of Stat. 1982. V. 10, № 4. P. 1269–1276.
478. *Нискина Н. П., Тейман А. И., Шмерлинг Л. С.* Непараметрические методы статистики, основанные на рангах, их применение. — Препринт / ВНИИ системных исследований. — М., 1986.

479. *Ansari A. R., Bradley R. A.* Rank-tests for dispersions // AMS. 1960. V. 31, № 4. P. 1174–1189.
480. *Siegel S., Tukey J. W.* A nonparametric sum of ranks procedure for relative spread in unpaired samples // JASA. 1960. V. 55, № 291. P. 429–445.
481. *Capon J.* Asymptotic efficiency of certain locally most powerfull rank tests // AMS. 1961. V. 32, № 1. P. 88–100.
482. *Šidak Z.* Tables for two normal-scores tests for the two-sample scale problem // Aplik. Matematiky. 1973. V. 18, № 5. P. 346–363.
483. *Klotz J.* Nonparametric tests for scale // AMS. 1962. V. 33, P. 498–512.
484. *Savage I. R.* Contributions to the theory of rank order statistics — the two-sample case // AMS. 1956. V. 27. P. 590–615.
485. *Šidak Z.* Tables for the two-sample Savage rank test optimal for exponential densities // Aplik. Mat. 1973. V. 18, № 5. P. 364–374.
486. *Sukhatme B. V.* On certain two-sample nonparametric tests for variances // AMS. 1957. V. 28, № 1. P. 188–194.
487. *Mood A.* On the asymptotic efficiency of certain nonparametric tests // AMS. 1954. V. 25. P. 514–522.
488. *Laubsher N. F., Steffenc F. E., De Lange E. M.* Exast critical values for Mood's distribution-free test statistic for dispersion and its normal approximation // Technometrics. 1968. V. 10, № 3. P. 497–507.
489. *Welchser S.* Mood's test for dispersion as a counting of triplets // Commun. Stat.-Theor. Meth. 1985. V. 14, № 2. P. 303–306.
490. *Sukhatme B. V.* A two-sample distribution-free test for comparing variances // Biometrika. 1958. V. 45. P. 544–548.
491. *Sukhatme B. V.* Testing the hypothesis that two population differ only in location // AMS. 1958. V. 29, № 1. P. 60–78.
492. *Sukhatme B. V.* On certain two-sample nonparametric tests for variances // AMS. 1957. V. 28, № 1. P. 188–199.
493. *Sandvik L., Olsson B.* A nearly distribution-free test for comparing dispersion in paired samples // Biometrika. 1982. V. 69, 32. P. 484–485.
494. *Krauth J., Lienert G. A.* Ein lokationsinsensitiver Dispersiontest für zwei unabhängige Stichproben [DP-test] // Biometr. Z. 1974. V. 16, № 2. S. 83–90.
495. *Krauth J., Lienert G. A.* Ein lokationsinsensitiver Dispersiontest für zwei abhängige Stichproben // Biometr. Z. 1974. V. 16, № 2. S. 91–95.
496. *Kamat A. R.* A two-sample distribution-free test // Biometrika. 1956. V. 43. P. 377–387.
497. *Bush J. R., Wieand H. S.* An asymptotically optimal nonparametric statistic for testing equally of two normal population means and variances // Commun. Stat.-Theor. Meth. 1982. V. 11, № 1. P. 1–12.
498. *Bhapkar V. P., Deshpande J. V.* Some nonparametric tests for multisample problems // Technometrics. 1968. V. 10, № 3. P. 578–585.
499. *Bhapkar V. P.* A nonparametric test for the problem of several samples // AMS. 1961. V. 32. P. 1108–1117.
500. *Ljung G. M., Box G. E. P.* On a measure of lack of fit in time-series models // Biometrika. 1978. V. 65. P. 197–203.
501. *Dufor J.-M., Roy R.* Some robust exact results on sample for randomness // J. of Econometrics. 1985. V. 29. P. 257–273.
502. *Bhapkar V. P.* A nonparametric test for the several sample location problem, Univ. of North Carolina, Inst. of Statistics, Mimeo series, № 411, 1964.
503. *Foster F. G., Stuart A.* Distribution-free tests in timeseries dated on the breaking of records // JRSS. 1954. V. B16, № 1. P. 1–22.
504. *Cox D. R., Stuart A.* Quick sing tests for trend in location and dispersion // Biometrika. 1955. V. 42. P. 80–95.
505. *Hsu D. A.* Tests for variance shift at an unknown time point // Appl. Statist. 1977. V. 26, № 3. P. 279–284.

506. *Hsieh H. K.* Nonparametric tests for scale shift at a unknown time point // Commun. Stat.-Theor. Meth. 1984. V. 13, № 11. P. 1335–1355.
507. *Wald A.*, *Wolfowitz J.* On a test whether two samples are from the some populations // AMS. 1940. V. 11. P. 147–162.
508. *Ramachandran G.*, *Ranganathan J.* // J. Madras Univ. 1953. Sec. B. V. 8. P. 76. (Цит. по [14]).
509. *Olmsted P. S.* Runs determined in a sample by an arbitrary cut // Bell. System. Technical J. 1958. V. 37. P. 55–82.
510. *Shaughnessy P. W.* Multiple runs distributions: Recurrence and critical values // JASA. 1981. V. 76, № 375. P. 732–736.
511. *Anderson R. L.* Distribution of the serial correlation coefficient // AMS. 1942. V. 13. P. 34–43.
512. *Knoke J. D.* Testing for randomness against autocorrelation: Alternative tests // Biometrika. 1977. V. 64, № 3. P. 523–529.
513. *Knoke J. D.* Testing for randomness against autocorrelation: The parametric case // Biometrika. 1975. V. 62. P. 571–575.
514. *Wald A.*, *Wolfowitz J.* An exact test for randomness in the non-parametric case based on serial correlation // AMS. 1943. V. 14. P. 378–388.
515. *Stuart A.* The efficient of tests on randomness against normal regression // JASA. 1956. V. 51. P. 285–287.
516. *Bartels R.* The rank version of von Neimann's ratio test for randomness // JASA. 1982. V. 77, № 377. P. 40–46.
517. *Mc Gielchrist C. A.*, *Woodyer K. D.* Note on a distribution-free CISIM technique // Technometrics. 1975. V. 17, № 3. P. 321–325.
518. *Woodward R. H.*, *Goldsmith P. L.* Cumulative sum techniques, ICI Monograph., № 3, Oliver and Boyd, London, 1964.
519. *Шенк Х.* Теория инженерного эксперимента / Пер. с англ. — М.: Мир, 1972.
520. *Уорсинг Ф.*, *Геффнер Л.* Методы обработки экспериментальных данных / Пер. с англ. — М.: ИЛ, 1949.
521. *Irwin J. O.* On a criterion for the rejection of outlying observations // Biometrika. 1925. V. 17. P. 238–250.
522. *Grubbs F. E.* Simple criteria for testing rejection of outlying observations // AMS. 1950. V. 21. P. 27–58.
523. *David H. A.* Revised upper percentage points of the extreme studentized deviate from the sample mean // Biometrika. 1956. V. 43. P. 450–461.
524. *Dixon W. J.* Rations involving extreme values // AMS. 1956. V. 22. P. 68–76.
525. *Hoaglin D. C.*, *Iglewicz B.* Fine-tuning some resistant rules for outlier labeling // JASA. 1987. V. 82, № 400. P. 1147–1149.
526. *Tietjen G.*, *Moore H.* Some Grubb's type statistics for the detection of several outliers // Technometrics. 1972. V. 14. P. 583–597.
527. *Смоляк С. А.*, *Титаренко Б. П.* Устойчивые методы оценивания. — М.: Статистика, 1980.
528. *Rosner B.* On the detection of many outliers // Technometrics. 1975. V. 17. P. 221–227.
529. *Rosner B.* Percentage points for the RST many outlier procedure // Technometrics. 1977. V. 19, № 3. P. 307–312.
530. Goodness-of-fit techniques, ed. D'Agostino R. B. and Stephens M. A., Marcel Dekker Inc., New York and Basel, 1986.
531. *Бродский Я. С.*, *Быцань Н. Н.*, *Власенко В. М.* Об исключении экстремальных значений // Завод. лаб. 1975. № 7. С. 847–849.
532. *Kimber A. C.* Tests for many outlier on a exponential sample // JRSS. Sec. C; Appl. Stat. 1982. V. 31. № 3. P. 263–271.
533. *Fung Karen Yuen*, *Paul S. R.* Comparisons of outlier detection procedures in Weibull or extreme-value distributions // Commun. Stat.-Simul. Comput. 1985. V. 14, № 4. P. 895–917.

534. *Mann N. R.* Optimal outlier tests for a Weibull model–to identify process changes or predict failure times // Studies in the Management Sciences. 1982. V. 19. P. 261–270.
535. *Bartlett V.*, *Lewis T.* Outlier in statistical data // N. Y.: J. Willey, 1978.
536. *Dixon R. J.* Analysis of extreme values // AMS. 1950. V. 21. P. 488–506.
537. *Beckman R. J.*, *Cook R. D.* „Outlier…s" // Technometrics. 1983. V. 25. P. 119–163.
538. *Darling D. A.* On a test for homogeneity and extreme values // AMS. 1952. V. 23, № 3. P. 450–456.
539. *Аронов И. З.*, *Золотарев А. О.* К вопросу оценки однородности экспериментальных данных об отказах изделий // Надежность и контроль качества. 1987. № 7. С. 3–6.
540. *Бахмутов В. Ф.*, *Аронов И. З.* Общий критерий оценки однородности и методика обработки результатов усталостных испытаний // Надежность и контроль качества. 1986. № 5. С. 25–29.
541. *Bowker A. H.* Computation of factors for tolerance limits on a normal distribution when sample is large // AMS. 1946. V. 17. P. 238–240.
542. *Stange K.* Angewandte Statistik, 1 und 2, Berlin–Heidelberg–New York, 1970/1971.
543. *Frawwley W. H.*, *Kapadia C. H.*, *Rao J. N.*, *Owen D. B.* Tolerance limits based on range ang mean range // Technometrics. 1971. V. 13. P. 651–655.
544. *Mitra S. K.* Tables for tolerance limits for normal population based on sample mean or mean range // JASA. 1957. V. 52. P. 88–94.
545. *Resnikoff G. L.* Two-sided tolerance limits for normal distribution using the range, Appl. Mathem. And Statistics Laboratory, Stanford Univ. Technical Report, № 33, 1957.
546. *Tietjen G. L.*, *Johnson M. E.* Exact statistical tolerance limits for sample variances // Technometrics. 1979. V. 21, № 1. P. 107–110.
547. *Wilcs S. S.* Statistical prediction with special reference to the problem of tolerance limits // AMS. 1942. V. 13. P. 400–409.
548. *Hall I. J.*, *Prairie R. R.* One-sided prediction interval to contain at least m out of k future observations // Technometrics. 1973. V. 15, № 4. P. 897–914.
549. *Lawless J. F.* A prediction problem concerning samples from the exponential distribution, with application in life testing // Technometrics. 1971. V. 13, № 4. P. 725–730.
550. *Hahn G. J.* A simultaneous prediction limits on the means of future samples from an exponential distribution // Technometrics. 1975. V. 17, № 3. P. 341–345.
551. *Moran P. A. P.* Some theorems on time series 2: The significance of the serial correlation coefficient // Biometrika. 1948. V. 35. P. 255–260.
552. *Hollin M.*, *Ingeubleek J.-F.*, *Puri M. L.* Linear serial rank tests for randomness against ARMA alternatives // Ann. Statist. 1985. V. 13. P. 1156–1181.
553. *Hollin M.*, *Ingenbleek J.-F.*, *Puri M. L.* Linear and quadratic serial rank tests for randomness against social dependence // J. Of Time Serial Annal. 1987. V. 8. P. 409–424.
554. *Hollin M.*, *Merald G.* Rank-tests for randomness against first order serial dependence // JASA. 1988. V. 83. P. 1117–1129.
555. *Hollin M.*, *Puri M. L.* Optimal rank-based procedures for time series analysis: Testing on ARMA model against other ARMA models // Annal. Statist. 1988. V. 16. P. 402–432.
556. *Hollin M.*, *Laforet A.*, *Merald G.* Distribution-free tests against dependence: signed or unsigned ranks? // J. of Stat. Planning and Inference. 1990. V. 24. P. 151–165.
557. *Mee R. W.* Simultaneous tolerance intervals for normal population with common variance // Technometrics. 1990. V. 32, № 1. P. 83–92.
558. *Odeh R. E.*, *Chou Youn-Min*, *Owen D. H.* Sample-size determination for two-sided β-expectation tolerance intervals for a normal distribution // Technometrics. 1989. V. 31, № 4. P. 461–468.
559. *Браунли К. А.* Статистические исследования в производстве. — М.: ИЛ, 1949.
560. *Нискина Н. П.*, *Тейман А. И.*, *Шмерлинг Л. С.* Двухфакторный дисперсионный анализ для неполных данных / Препринт. — М.: ВНИИСИ, 1986.
561. *Prentice M. J.* On the problem of m incomplete ranking // Biometrika. V. 66, P. 167–177.
562. *Mack D. A.*, *Skillings J. H.* A Friedman-type rank test for main effects in a two-way ANOVA // JASA. 1980. V. 75, 3372. P. 941–947.

563. *Mack D. A.* A quick and easy distribution-free test for main effects in a two-way ANOVA // Commun. Stat. 1981. V. 10, № 6. P. 571–591.
564. *Stange K.* Über einen zweiseitigen Test für die Korrelationszahle einer zweidimensionalen Normaverteilung // Statist. H. 1973. V. 14, № 3. S. 206–236.
565. *Cox D. R., Stuart A.* Quick tests for trend in location and dispersion // Biometrika. 1955. V. 42. P. 80–95.
566. *Nelson L. S.* A sign test for correlation // J. of Quality Technology. 1983. V. 15, № 4. P. 199–200.
567. *Elandt R. C.* Exact and approximate power function of the non-parametric test of tendency // AMS. 1962. V. 33, № 2. P. 471–481.
568. *Kruskal W. H.* Ordinal measures of association // JASA. 1958. V. 53. P. 814–861.
569. *Olmstead P. S., Tukey J. W.* A corner test for association // AMS. 1947. V. 18. P. 495–513.
570. *Blum J. R., Kiefer J., Rosenblatt M.* Distribution-free tests of independence based on the sample distribution function // AMS. 1961. V. 32. P. 485–498.
571. *Iman R. L., Conover W. J.* Approximation of the critical region for Spierman's rho with and without ties present // Commun. Stat.-Simul. And Comput. 1978. V. 7. P. 269–282.
572. *Hoeffding W.* A non-parametric test of independence // AMS. 1948. V. 19. P. 546–557.
573. *Shirahate S.* Intraclass rank tests for independence // Biometrika. 1981. V. 68, № 2. P. 451–456.
574. *Fieller E. C., Pearson E. S.* Tests for rank correlation coefficients. 2 // Biometrika. 1961. V. 48, № 1–2. P. 29–40.
575. *Kendall M. G., Babington Smith.* The problem of m rankings // AMS. 1939. V. 10. P. 275–287.
576. *Schucany W. R., Frawley W. H.* A rank test for two group concordance // Psichometrika. 1973. V. 38. P. 249–258.
577. *McNemar Q.* Note on sample error of the differences between correlated proportions or percentages // Psychometrika. 1947. V. 12. P. 153–154.
578. *Woolf B.* The log likelihood ratio test [the G-test]. Methods and tables for tests of heterogeneity in contingency tables // Ann. Human Genetics. 1957. V. 21. P. 397–409.
579. *Le Roy H. L.* Ein einfacher χ^2-Test für den Simultanvergleich der inneren Struktur von zwei Analogen 2×2, Haufigkeitstabellen mit freien Kolonnen und Zeilentotallen, Schweizer. Landw. Forchg., 1, 1962, S. 451–454.
580. *Kvetom K.* Formation of empirical regression curves and surfaces using power function // Acta Technica. 1988. № 2. P. 141–157.
581. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ / Пер с англ. — М.: Статистика, 1973.
582. *Йоала В., Ольман В.* Метод оценивания коэффициентов линейной одномерной регрессии // Изв. АН ЭССР: Физика. Математика. 1987. Т. 36, № 4. С. 422–424.
583. *Маслов П. П.* Замена способа наименьших квадратов простейшим расчетом // Вестник статистики. 1973. № 5. С. 18–20.
584. *Bartlett M. S.* Fitting a straight line when both variables are subject to error // 1949. Biometrics. V. 5. P. 207–212.
585. *Kerrich J. E.* Fitting the line $y = b \cdot x$ when errors of observation are present in both variables // The Amer. Statistician. 1966. V. 20. P. 24.
586. *Sen P. K.* Estimates of the regression coefficient based on Kendall's tau // JASA. 1968. V. 63, № 324. P. 258–264.
587. *Siegel A. F.* Robust regression using repeated medians // Biometrika. 1982. V. 69, № 1. P. 242–244.
588. *Daniels H. F. F* distribution-free test for regression parameters // AMS. 1953. V. 25, № 3. P. 499–513.
589. *Brown G. W., Mood A. M.* On median tests for linear hypotheses // Proc. of the Second Berkeley Symp. on Math. Stat. And Prob., Univ. of Calif. Press., 1950. P. 159–166.
590. *Theil H.* A rank-invariant method of linear and polynomial regression analysis. 1 // Proc. Kon. Ned. Akad. V. Weetenseh. 1950. A53. P. 386–392.

591. *Durbin J., Watson G. S.* Testing for serial correlation in least-squares regression // Biometrika. 1951. V. 38. P. 159–178.
592. *Acton F. S.* Analysis of straight line data. — N. Y.: Welly, 1959. — P. 261.
593. *Tietjen G. L., Moore R. H., Beckman R. J.* Testing for a single outlier in simple linear regression // Technometrics. 1973. V. 15, № 4. P. 717–721.
594. *Prescott P.* An approximate test for outlier in linear models // Technometrics. 1975. V. 17, № 1. P. 129–132.
595. *Lund R. E.* Tables for an approximate test for outliers in linear models // Technometrics. 1975. V. 17, № 4. P. 473–476.
596. *Wallis W. A.* Tolerance intervals for linear regression // Proc. of the Second Berkley Symp. on Math. Statist. And Probability, Univ. of Calif. Press., Berkley and Los Angeles, 1951. P. 43–51.
597. *Working H., Hotelling H.* Applications of the theory of error to the interpretation of trends // JASA. 1929. V. 24. P. 73–85.
598. *Пустыльник Е. И.* Две задачи, связанные с обработкой линейных зависимостей // Завод. лаб. 1981. Т. 47, № 3. С. 52–53.
599. *Янкаускас В.* Модель регрессии с автокоррелированными остатками в условиях малых выборок // Применение теории вероятностей и математической статистики: Сб. 1983. Вып. 5. Вильнюс, 1983. — С. 86–94.
600. *Четыркин Е. М.* Статистические методы прогнозирования. — М.: Статистика, 1975.
601. *Петрович С. Л., Шлеч Г. К.* Робастная регрессия: оценки и сравнения методом Монте-Карло // Завод. лаб. 1987. № 3. С. 41–48.
602. *Айвазян С. А., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика. Исследование зависимостей. — М.: Финансы и статистика, 1985.
603. *Демиденко Е. З.* Линейная и нелинейная регрессия. — М.: Статистика, 1981.
604. *Мудров В. И., Кушко В. Л.* Метод наименьших модулей. — М.: Знание, 1971.
605. *Forsythe A. B.* Robust estimation of straight line regression coefficients by minimizing p-th power deviations // Technometrics. 1972. V. 14. P. 159–166.
606. *Епишин Ю. Г.* Регрессионный метод наименьших абсолютных отклонений // Завод. лаб. 1974. № 10. С. 1227–1232.
607. *Епишин Ю. Г.* Об оценках параметров наименьших абсолютных отклонений // Экономика и математические методы. 1974. Т. X, № 54. С. 1023. Р. 1028.
608. *Дронов В. С., Целищев В. Д.* О сглаживании экспериментальных зависимостей методом однозначной аппроксимации // Завод. лаб. 1975. Т. 41, № 7. С. 844–846.
609. Планирование эксперимента в исследовании технологических процессов / Хартман К. и др. — Под ред. Э. К. Лецкого. — М.: Мир, 1977.
610. *Шиндовский Э., Щюрц О.* Статистические методы управления качеством. — М.: Наука, 1964.
611. *Джонсон Н., Мюн Ф.* Статистика и планирование эксперимента в технике и науке. Методы обработки данных / Пер. с англ. — М.: Мир, 1980.
612. *Shewhart W. A.* Economic control of quality of manufactured product, van Nostrond, New York, 1931.
613. *Barnard G. A.* Cumulative charts and stochastic processes // JRSS. Ser. B. 1959. V. 21. P. 148–159.
614. *Goldsmit P. L., Whitfield H.* Average run lenght in cumulative sum chart when a V-mask is used // Technometrics. 1961. V. 3. P. 441–447.
615. *Johnson N. L., Leone F. C.* Cumulative sum control charts // Industr. Quality Contr. 1962. V. 18–19 [three papers].
616. *Page E. S.* Cumulative sum charts // Technometrics. 1961. V. 3. P. 511–518.
617. *Cox D. R.* The use of range in sequential analysis // JRSS. 1949. Ser. B, V. 11. P. 921–929.
618. *Shainin D.* How to improve upon the benefits of the good old and control charts // Proc. World Quality Congress, Braiton, 1984, V. 2. P. 48–56.
619. *Налимов В. В.* Планирование эксперимента. — М.: Наука, 1971.
620. *Налимов В. В., Голикова Т. И.* Логические основания планирования эксперимента. — М.: Изд-во МГУ, 1971.

621. *Налимов В. В., Чернова Н. А.* Статистические методы планирования экстремальных экспериментов. — М.: Наука, 1965.
622. *Адлер Ю. П.* Введение в планирование эксперимента. — М.: Металлургия, 1969.
623. *Финни Д. Д.* Введение в теорию планирования эксперимента. — М.: Наука, 1970.
624. *Федоров В. В.* Теория оптимального эксперимента (планирование регрессионных экспериментов). — М.: Наука, 1971.
625. *Хикс Ч. Р.* Основные принципы планирования эксперимента. — М.: Мир, 1967.
626. *Box G. E. P., Wilson K. B.* On the experimental attainment of optimum conditions // JRSS. 1951. V. B13, №1. P. 1–17.
627. *Box G. E. P., Hunter J. S.* Condensed calculation for evolutionary operation programs // Technometrics. 1959. V. 1. P. 77–84.
628. *Kiefer J.* // JRSS. 1959. V. B21. P. 272.
629. *Голикова Т. И., Микешина Н. Г.* Свойства D-оптимальных планов и методы их построения // Новые идеи в планировании эксперимента: Сб. — М.: Наука, 1969. — С. 21–58.
630. *Андрукович П. Ф., Голикова Т. И., Костина С. Г.* Планы второго порядка на гиперкубе, близкие по свойствам к D-оптимальным // Новые идеи в планировании экспериментов: Сб. — М.: Наука, 1969. — С. 140–153.
631. *Hartey H. O.* Smallest composite design for quadratic response surfase // Biometrics. 1959. V. 15. P. 611–622.
632. *Адлер Ю. П., Маркова Е. В., Грановский Ю. В.* Планирование эксперимента при поиске оптимальных условий. — М.: Наука, 1970.
633. *Уайльд Д. Дж.* Методы поиска экстремума / Пер. с англ. — М.: Наука, 1967.
634. *Spendley W., Hext G. R., Himsworth F. R.* Sequential application of simplex design in optimization and evolutionary operations // Technometrics. 1962. V. 4. P. 441.
635. *Копаев Б. В.* К вопросу использования относительных разностей в методе наименьших квадратов // Геодезия и картография. 1991. №6. С. 9–13.
636. *Копаев Б. В.* Использование условно-относительных разностей при определении коэффициентов полинома методом наименьших квадратов // Надежность и контроль качества. 1991. №12. С. 3–12.
637. *Штефан В. В., Карташев В. Я.* Исследование статистических зависимостей методом обратных разделенных разностей. — Препринт / ВИНИТИ. №1105-В 92. — С. 2–19.
638. *Абдуллаева В. Г., Володченко А. П.* Определение параметров уравнения регрессии в форме нормального уравнения прямой. — Препринт / ВИНИТИ. №1246–1989. 1989. — С. 2–11.

Сокращенные названия использованных журналов

AMS — The Annals of Mathematical Statistics
JRSS — Journal of the Royal Statistical Society
JASA — Journal of the American Statistical Association
CS — Communication in Statistics
JSCS — Journal of Statistical Computation and Simulation
RISI — Review of the International Statistical Institute
SJS — Scandinavian Journal of Statistics
SASJ — South African Statistical Journal
SN — Statistical Neerlandica
JIS — Journal of the Institute of Statisticians
AS — The American Statistican
AISM — Annals of the Institute of Statistical Mathematics
CJS — The Canadian Journal of Statistics

Перечень демонстрационных задач

№	Стр.	Исходная информация	Содержание задачи
1. Вычисление вероятностей для полностью определенных законов распределения вероятностей			
1	28	Нормальное распределение	Вычисление значения функции распределения
2	30	Нормальное распределение	Вычисление квантилей
3	31	Нормальное распределение	Вычисление математических ожиданий порядковых статистик
4	32	Нормальное распределение	Комплексное вычисление вероятностных характеристик
5	34	Равномерное распределение	Комплексное вычисление вероятностных характеристик
6	35	Логарифмически нормальное распределение	Вычисление значения функции распределения
7	36	Экспоненциальное распределение	Вычисление вероятностных характеристик применительно к испытаниям приборов
8	38	Распределение Вейбулла	Вычисление вероятностных характеристик
9	39	Гамма-распределение	Вычисление вероятностных характеристик
10	43	Бета-распределение	Вычисление квантилей распределения
11	43	Бета-распределение	Вычисление квантилей распределения
12	44	Бета-распределение	Вычисление квантилей распределения
13	48	χ^2-распределение (распределение Пирсона)	Вычисление квантилей распределения
14	54	Распределение Стьюдента (t-распределение)	Вычисление квантилей распределения
15	59	Распределение Фишера (F-распределение)	Вычисление квантилей распределения
16	60	Распределение Фишера (F-распределение)	Вычисление квантилей распределения
17	62	Усеченное нормальное распределение	Вычисление изменения параметров при усечении нормального распределения

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
18	63	Распределение модуля случайной величины, распределенной нормально	Вычисление вероятностей и параметров распределения
19	64	Распределение, порождаемое нормальной плотностью с линейным дрейфом среднего	Вычисление вероятностных характеристик
20	66	Распределение, порождаемое нормальной плотностью с линейным дрейфом среднеквадратического отклонения	Вычисление вероятностных характеристик
21	68	Распределение Рэлея	Вычисление функции распределения и квантилей
22	69	Распределение Максвелла	Вычисление функции распределения
23	71	Распределение экстремального значения	Вычисление функции распределения и квантилей
24	72	Треугольное распределение (распределение Симпсона)	Вычисление функции распределения
25	72	Распределение Коши	Вычисление вероятностных характеристик
26	73	Логистическое распределение	Вычисление функции распределения и квантилей
27	74	Распределение Парето	Вычисление вероятностных характеристик
28	78	Композиции законов распределения вероятностей случайных величин, возникающие при расчете надежности по схеме „нагрузка–напряжение“	Вычисление вероятности безотказной работы, когда прочность и нагрузка распределены нормально
29	78	Композиции законов распределения вероятностей случайных величин, возникающие при расчете надежности по схеме „нагрузка–напряжение“	Вычисление вероятности безотказной работы, когда прочность и нагрузка распределены логарифмически нормально
30	79	Композиции законов распределения вероятностей случайных величин, возникающие при расчете надежности по схеме „нагрузка–напряжение“	Вычисление вероятности безотказной работы, когда прочность и нагрузка распределены экспоненциально
31	79	Композиции законов распределения вероятностей случайных величин, возникающие при расчете надежности по схеме „нагрузка–напряжение“	Вычисление вероятности безотказной работы, когда прочность распределена нормально, нагрузка — экспоненциально
32	79	Композиции законов распределения вероятностей случайных величин, возникающие при расчете надежности по схеме „нагрузка–напряжение“	Вычисление вероятности безотказной работы, когда прочность имеет гамма-распределение, нагрузка распределена экспоненциально

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
33	79	Композиции законов распределения вероятностей случайных величин, возникающие при расчете надежности по схеме „нагрузка–напряжение“	Вычисление вероятности безотказной работы, когда прочность имеет распределение Вейбулла, нагрузка распределена нормально
34	80	Нецентральное распределение Стьюдента (нецентральное t-распределение)	Вычисление параметра нецентральности
35	81	Нецентральное распределение Пирсона (нецентральное хи-квадрат распределение)	Вычисление квантили распределения
36	82	Нецентральное распределение Фишера (нецентральное F-распределение)	Вычисление квантили распределения
37	86	Биномиальное распределение (распределение Бернулли)	Вычисление функции распределения
38	88	Биномиальное распределение (распределение Бернулли)	Вычисление функции распределения аппроксимацией Моленара
39	89	Распределение Пуассона	Вычисление функции распределения
40	91	Отрицательное биномиальное распределение	Вычисление функции распределения
41	91	Распределение Паскаля	Вычисление функции распределения
42	92	Геометрическое распределение (распределение Фарри)	Вычисление плотности и функции распределения
43	94	Гипергеометрическое распределение	Вычисление функции распределения
2. Оценка параметров распределения вероятностей			
44	103	Нормальное распределение	Точечная оценка среднего значения
45	109	Нормальное распределение	Интервальная оценка среднего значения при неизвестной дисперсии
46	116	Нормальное распределение	Точечная оценка дисперсии и стандартного отклонения
47	121	Нормальное распределение	Интервальная оценка стандартного отклонения
48	131	Нормальное распределение (усеченная выборка)	Оценка параметров распределения
49	131	Нормальное распределение (неполностью определенная выборка)	Оценка параметров распределения
50	132	Нормальное распределение (цензурированная выборка)	Оценка параметров распределения
51	137	Экспоненциальное распределение	Точечная оценка интенсивности отказов при испытании партии приборов до заданной наработки с заменой отказавших приборов

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
52	137	Экспоненциальное распределение	Точечная оценка интенсивности отказов при испытании партии приборов до заданного числа отказов с заменой отказавших приборов
53	137	Экспоненциальное распределение	Точечная оценка интенсивности отказов при испытании партии приборов до заданной наработки без замены отказавших приборов
54	137	Экспоненциальное распределение	Точечная оценка интенсивности отказов при испытании партии приборов до заданного числа отказов без замены отказавших приборов
55	139	Экспоненциальное распределение	Двухстадийная оценка среднего значения
56	139	Экспоненциальное распределение (цензурированная выборка)	Точечная оценка параметра (среднего значения) с помощью порядковых статистик
57	144	Экспоненциальное распределение	Интервальная оценка среднего значения
58	144	Экспоненциальное распределение	Интервальная оценка интенсивности отказов при испытании партии приборов до заданной наработки с заменой отказавших приборов
59	144	Экспоненциальное распределение	Интервальная оценка интенсивности отказов при испытании партии приборов до заданного количества отказов с заменой отказавших приборов
60	144	Экспоненциальное распределение	Интервальная оценка интенсивности отказов при испытании партии приборов до заданного количества отказов без замены отказавших приборов
61	145	Экспоненциальное распределение	Интервальная оценка интенсивности отказов при испытании партии приборов до заданной наработки без замены отказавших приборов
62	145	Экспоненциальное распределение	Интервальная оценка наработки на отказ
63	173	Распределение Вейбулла	Точечная оценка параметра α при известном параметре β
64	173	Распределение Вейбулла	Совместная точечная оценка параметров распределения
65	174	Распределение Вейбулла	Совместная точечная оценка параметров распределения (метод моментов)
66	175	Распределение Вейбулла	Совместная точечная оценка параметров распределения по одной порядковой статистике

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
67	175	Распределение Вейбулла	Точечная оценка параметров распределения методом наименьших квадратов
68	176	Распределение Вейбулла	Точечная оценка параметров распределения методом аппроксимации
69	177	Распределение Вейбулла	Точечная оценка параметров распределения
70	177	Распределение Вейбулла (выборка цензурирована)	Точечная оценка параметров распределения
71	177	Распределение Вейбулла	Интервальная оценка параметра α при известном параметре β
72	178	Распределение Вейбулла	Совместная интервальная оценка параметров распределения
73	178	Распределение Вейбулла (выборка цензурирована)	Интервальная оценка квантили распределения
74	178	Распределение Вейбулла	Интервальная оценка квантили распределения
75	178	Экспоненциальное распределение. Распределение Вейбулла с известным параметром β (выборка цензурирована)	Оценка экономии в продолжительности испытаний приборов при цензурировании выборов
76	181	Гамма-распределение	Точечная оценка параметров α и β. Интервальная оценка параметра β при известном параметре α
77	187	Биномиальное распределение	Интервальная оценка параметра распределения
78	191	Гипергеометрическое распределение	Оценка верхней границы количества дефектных изделий изделий в партии
79	193	Закон распределения вероятностей неизвестен (известна только дисперсия)	Интервальная оценка центра распределения
80	194	Закон распределения вероятностей неизвестен (известна только дисперсия)	Оценка квантили распределения
81	195	Экспоненциальное распределение с изменяющейся интенсивностью отказов	Оценка интенсивности отказов изделия в период приработки
82	196	Экспоненциальное распределение	Прогнозирование средней наработки для будущих испытаний
3. Определение объема выборки, необходимого для оценки параметра распределения с заданной точностью			
83	197	Нормальное распределение (дисперсия известна)	Определение объема выборки, необходимого для оценки среднего значения с заданной абсолютной ошибкой

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
84	198	Нормальное распределение (дисперсия неизвестна)	Определение объема выборки, необходимого для оценки среднего значения с заданной предельной относительной ошибкой
85	198	Распределение Вейбулла (параметр формы β известен)	Определение объема выборки, необходимого для оценки среднего значения с заданной относительной погрешностью
86	200	Биномиальное распределение	Определение объема выборки, необходимого для контроля заданной доли дефектных изделий в партии
87	200	Биномиальное распределение	Определение объема выборки, обеспечивающего заданную длину доверительного интервала для оценки параметра распределения
88	200	Биномиальное распределение	Определение объема выборки, необходимого для прогнозирования числа дефектных изделий в партии
89	201	Экспоненциальное распределение	Определение объема выборки для оценки интенсивности отказов
90	201	Гамма-распределение	Определение объема выборки, необходимого для оценки среднего значения с заданной предельной ошибкой
4. Проверка согласия эмпирического распределения с предполагаемым теоретическим законом			
91	207	Распределение вероятностей неизвестно (одна выборка)	Проверка критерием χ^2 гипотезы о согласии эмпирического распределения с нормальным законом
92	211	Распределение вероятностей неизвестно (две выборки)	Проверка гипотезы согласия эмпирического распределения с экспоненциальным с помощью критерия числа пустых ячеек. Проверка совпадения распределений в двух выборках
93	212	Распределение вероятностей неизвестно (две выборки)	Проверка совпадений эмпирических распределений в двух выборках критерием Барнетта–Эйсена
94	215	Распределение вероятностей неизвестно (одна выборка)	Проверка совпадения эмпирического распределения с нормальным (при заданных параметрах) критерием Колмогорова–Смирнова
95	217	Распределение вероятностей неизвестно (одна выборка)	Проверка совпадения эмпирического распределения с нормальным законом критерием Смирнова–Крамера–фон Мизеса
96	219	Распределение вероятностей неизвестно (одна выборка)	Проверка совпадения эмпирического распределения с нормальным законом критерием Реньи

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
97	221	Распределение вероятностей неизвестно (одна выборка)	Проверка совпадения эмпирического распределения с нормальным законом критерием Андерсона–Дарлинга
98	223	Распределение вероятностей неизвестно (одна выборка)	Проверка совпадения эмпирического распределения с нормальным законом критерием Ватсона
99	224	Распределение вероятностей неизвестно (одна выборка)	Проверка совпадения эмпирического распределения с нормальным законом критерием Купера
100	226	Распределение вероятностей неизвестно (одна выборка)	Проверка совпадения эмпирического распределения с нормальным законом критерием Дарбина
5. Проверка совпадения функций распределения вероятностей в двух выборках			
101	228	Распределение вероятностей неизвестно	Проверка совпадения эмпирических распределений критерием Колмогорова–Смирнова
102	229	Распределение вероятностей неизвестно	Проверка совпадения эмпирических распределений критерием Катценбайссера–Хакля
103	230	Распределение вероятностей неизвестно	Проверка совпадения эмпирических распределений критерием Андерсона
6. Проверка совпадения эмпирического распределения вероятностей с нормальным законом			
104	232	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием χ^2
105	234	Распределение вероятностей неизвестно	Проверка нормальности критериями Колмогорова–Смирнова и $n\omega^2$ с оценкой параметров гипотетического распределения по выборке
106	235	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Фроцини
107	239	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Шапиро–Уилка
108	242	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Васичека
109	243	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Хегази–Грина
110	245	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Али–Чёрго–Ревеса

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
111	246	Распределение вероятностей неизвестно (одна выборка)	Проверка нормальности эмпирического распределения критерием Филлибена
112	252	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Ла Брека
113	253	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Локка–Спурье
114	256	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Оя
115	258	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Гири
116	260	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Дэвида–Хартли–Пирсона
117	260	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Шпигельхальтера
118	262	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Саркади
119	264	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Лина–Мудхолкара
120	265	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Мартинеса–Иглевича
121	268	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Д'Агостино
122	270	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критериями асимметрии и эксцесса
123	272	Распределение вероятностей неизвестно	Проверка нормальности эмпирического распределения критерием Муроты–Такеучи
124	276	Распределение вероятностей неизвестно (совокупность выборок малого объема)	Проверка нормальности эмпирического распределения по совокупности независимых выборок малого объема
7. Проверка совпадения эмпирического распределения с экспоненциальным законом			
125	281	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Шапиро–Уилка

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
126	282	Распределение вероятностей неизвестно (выборка цензурирована)	Проверка экспоненциальности эмпирического распределения критерием Шапиро–Уилка
127	284	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерями типа Колмогорова–Смирнова
128	287	Распределение вероятностей неизвестно (выборка цензурирована)	Проверка экспоненциальности эмпирического распределения критериями типа Смирнова–Крамера–фон Мизеса и Колмогорова–Смирнова
129	288	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Фроцини
130	289	Распределение вероятностей неизвестно	Проверка экспоненциальности распределения корреляционным критерием
131	291	Распределение вероятностей неизвестно (выборка цензурирована)	Проверка экспоненциальности эмпирического распределения критерием Брейна–Шапиро
132	293	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Кимбера–Мичела
133	293	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Фишера
134	294	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Бартлетта–Морана
135	295	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Климко–Антла–Радемакера–Рокетта
136	297	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Холлендера–Прошана
137	299	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Кочара
138	301	Распределение вероятностей неизвестно	Проверка экспоненциальности распределения критерием Эппса–Палли
139	302	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Бергмана
140	304	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Шермана

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
141	304	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием наибольшего интервала
142	305	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Хартли
143	306	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием экспоненциальных меток
144	307	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критериями независимости интервалов
145	310	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критериями $\tilde{U}_n$ Гринвуда
146	311	Распределение вероятностей неизвестно	Проверка гипотезы согласия эмпирического распределения с двухпараметрическим распределением Вейбулла (частный случай экспоненциального распределения) критерием Манна–Фертига–Шуера
147	316	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Дешпанде
148	317	Распределение вероятностей неизвестно	Проверка экспоненциальности эмпирического распределения критерием Лоулесса
8. Проверка совпадения эмпирического распределения с равномерным законом			
149	319	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критерием Шермана
150	322	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критерием Морана
151	323	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критерием Ченга–Спиринга
152	324	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критерием Саркади–Косика
153	326	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критерием Дудевича–ван дер Мюлена
154	327	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критерием Хегази–Грина

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
155	329	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критерием Янга
156	330	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критериями типа Колмогорова–Смирнова
157	331	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критерием Фроцини
158	333	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения критерием Гринвуда–Кэсенберри–Миллера
159	333	Распределение вероятностей неизвестно	Проверка равномерности эмпирического распределения „сглаженным“ критерием Неймана–Бартона
9. Проверка симметричности эмпирического распределения вероятностей			
160	336	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения „быстрым“ критерием Кенуя
161	337	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения критерием Смирнова
162	339	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения критерием знаков
163	340	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения критерием Вилкоксона
164	341	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения критерием знаков Антилла–Керстинга–Цуккини
165	343	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения критерием Бхатачарья–Гаствирта–Райта
166	345	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения критерием Финча
167	346	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения критерием Бооса
168	349	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения критерием Гупты
169	351	Распределение вероятностей неизвестно	Проверка симметричности эмпирического распределения критерием Фрезера

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
10. Подбор кривых распределения вероятностей по эмпирическим данным			
170	353	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой из семейства S_L Джонсона
171	356	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой из семейства S_B Джонсона
172	366	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой из семейства S_U Джонсона
173	373	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой Пирсона типа I
174	375	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой Пирсона типа II
175	377	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой Пирсона типа III
176	379	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой Пирсона типа IV
177	380	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой Пирсона типа V
178	381	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой Пирсона типа VI
179	383	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения кривой Пирсона типа VII
180	384	Распределение вероятностей неизвестно	Аппроксимация эмпирического распределения с помощью разложения нормальной плотности распределения
181	386	Распределение вероятностей неизвестно	Оценка плотности эмпирического распределения методом бета-вкладов
11. Сравнение параметров нормально распределенных совокупностей			
182	390	Нормальное распределение (две выборки с известными дисперсиями)	Проверка гипотезы равенства средних в двух совокупностях
183	390	Нормальное распределение (две выборки с неизвестными, но равными дисперсиями)	Проверка гипотезы равенства средних в двух совокупностях
184	392	Нормальное распределение (две выборки с неизвестными и неравными дисперсиями)	Проверка гипотезы равенства средних в двух совокупностях
185	393	Нормальное распределение (две выборки)	Проверка гипотезы равенства средних в двух совокупностях модифицированным критерием Стьюдента
186	393	Нормальное распределение (две парные выборки)	Проверка гипотезы равенства средних в двух совокупностях парным t-критерием

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
187	394	Нормальное распределение (одна выборка и одно независимое значение случайной величины)	Проверка гипотезы о принадлежности независимого выборочного значения к первичной выборке критерием Уолша
188	396	Нормальное распределение (две выборки)	Проверка гипотезы равенства средних в двух совокупностях двухступенчатым критерием Волфа
189	397	Нормальное распределение (две выборки с равными дисперсиями)	Проверка гипотезы равенства средних в двух совокупностях F-критерием
190	398	Нормальное распределение (более двух выборок с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях модифицированым критерием Стьюдента
191	399	Нормальное распределение (более двух выборок с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях критерием „стьюдентизированного“ размаха
192	401	Нормальное распределение (более двух выборок с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях дисперсионным критерием
193	402	Нормальное распределение (более двух выборок с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях критерием Полсона
194	403	Нормальное распределение (более двух выборок с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях методом прямого сравнения Тьюки
195	405	Нормальное распределение (более двух выборок с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях критерием „стьюдентизированного“ максимума (обобщенный критерий Тьюки)
196	407	Нормальное распределение (более двух выборок с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях критерием Шеффе
197	407	Нормальное распределение (более двух выборок равного объема с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях критерием Стьюдента–Ньюмена–Кейлса
198	408	Нормальное распределение (более двух выборок равного объема с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях критерием Дункана
199	412	Нормальное распределение (более двух выборок равного объема с равными дисперсиями)	Проверка гипотезы равенства средних в более, чем двух совокупностях критерием Линка–Уоллеса

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
200	413	Нормальное распределение (две выборки)	Проверка гипотезы равенства дисперсий в двух совокупностях критерием Фишера
201	413	Нормальное распределение (две выборки)	Проверка гипотезы равенства дисперсий в двух совокупностях критерием Романовского
202	414	Нормальное распределение (две выборки)	Проверка гипотезы равенства дисперсий в двух совокупностях критерием отношения размахов
203	415	Нормальное распределение (две выборки)	Проверка гипотезы равенства дисперсий в двух совокупностях критерием „стьюдентизированного" размаха
204	416	Нормальное распределение (две выборки)	Проверка гипотезы равенства дисперсий в двух совокупностях критерием Аризоно–Охты
205	417	Нормальное распределение (более двух выборок)	Проверка гипотезы равенства дисперсий в более, чем двух выборках критерием Бартлетта
206	418	Нормальное распределение (более двух выборок равного объема)	Проверка гипотезы равенства дисперсий в более, чем двух выборках критерием Кохрана
207	420	Нормальное распределение (более двух выборок равного объема)	Проверка гипотезы равенства дисперсий в более, чем двух выборках критерием Неймана–Пирсона
208	421	Нормальное распределение (более двух выборок равного объема)	Проверка гипотезы равенства дисперсий в более, чем двух выборках критерием Блисса–Кохрана–Тьюки
209	422	Нормальное распределение (более двух выборок равного объема)	Проверка гипотезы равенства дисперсий в более, чем двух выборках критерием Хартли
210	423	Нормальное распределение (более двух выборок равного объема)	Проверка гипотезы равенства дисперсий в более, чем двух выборках критерием Кэдуэлла–Лесли–Брауна
211	423	Нормальное распределение (более двух выборок)	Проверка гипотезы равенства дисперсий в более, чем двух выборках критерием Самиуддина
12. Сравнение параметров экспоненциально распределенных совокупностей			
212	425	Экспоненциальное распределение (две выборки)	Проверка гипотезы равенства средних в двух совокупностях критерием Фишера
213	425	Экспоненциальное распределение (две выборки)	Проверка равенства интенсивностей отказов критерием Фишера

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
214	426	Экспоненциальное распределение (две выборки)	Проверка равенства интенсивностей отказов в двух совокупностях двухвыборочным пуассоновским критерием
215	428	Экспоненциальное распределение (одна выборка и заданное значение интенсивности отказов)	Сравнение интенсивности отказов с заданным значением при испытании приборов до заданной наработки с заменой отказавших приборов
216	428	Экспоненциальное распределение (одна выборка и заданное значение интенсивности отказов)	Сравнение интенсивности отказов с заданным значением при испытании приборов до заданного количества отказов с заменой отказавших приборов
217	428	Экспоненциальное распределение (одна выборка и заданное значение интенсивности отказов)	Вариант решения задачи 216 без замены отказавших приборов
218	429	Экспоненциальное распределение (одна выборка и заданное значение интенсивности отказов)	Сравнение интенсивности отказов с заданным значением при испытании приборов до заданной наработки без замены отказавших приборов
219	429	Экспоненциальное распределение (одна выборка и заданное значение интенсивности отказов)	Сравнение интенсивности отказов с заданным значением при испытании приборов до заданного количества отказов без замены отказавших
220	430	Экспоненциальное распределение (более двух выборок равного объема)	Проверка гипотезы равенства интенсивностей отказов в более, чем двух совокупностях критерием Дэвида
221	430	Экспоненциальное распределение (более двух выборок). Фиксируются моменты наступления заданного количества отказов	Проверка гипотезы равенства интенсивностей отказов в более, чем двух совокупностях критерием максимального правдоподобия
222	431	Экспоненциальное распределение (более двух выборок). Фиксируются количества отказов в выборках за заданное время испытаний	Проверка гипотезы равенства интенсивностей отказов в более, чем двух совокупностях критерием максимального правдоподобия
223	431	Экспоненциальное распределение (более двух выборок равного объема)	Проверка гипотезы равенства интенсивностей отказов в более, чем двух совокупностях критерием отношения правдоподобия (критерий Нагарсенкера)
224	432	Экспоненциальное распределение (более двух выборок равного объема)	Проверка гипотезы равенства параметра положения (гарантийной наработки) в более, чем двух совокупностях критерием Чена

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
225	434	Экспоненциальное распределение (более двух выборок)	Проверка гипотезы равенства параметра положения (гарантийной наработки) и параметра масштаба (средней наработки, отсчитываемой от гарантийной наработки) в более, чем двух совокупностях комбинированным критерием Сингха
13. Сравнение параметров биномиально распределенных совокупностей			
226	436	Биномиальное распределение (две выборки)	Сравнение параметров распределений в двух совокупностях
227	437	Биномиальное распределение (одна выборка и заданное значение параметра)	Сравнение параметра распределения с заданным значением
228	438	Биномиальное распределение (более двух выборок)	Сравнение параметров распределения в более, чем двух совокупностях
14. Проверка гипотезы о значениях параметров распределения методом последовательного анализа Вальда			
229	445	Нормальное распределение с известным стандартным отклонением	Проверка гипотезы о значении среднего нормально распределенной совокупности. Вычисление параметров последовательной процедуры
230	447	Нормальное распределение с известным средним	Проверка гипотезы о значении дисперсии нормально распределенной совокупности. Вычисление параметров последовательной процедуры
231	448	Экспоненциальное распределение	Проверка гипотезы о значении параметра экспоненциально распределенной совокупности (средняя наработка на отказ). Вычисление параметров последовательной процедуры
232	450	Биномиальное распределение	Проверка гипотезы о значении параметра биномиально распределенной совокупности (доля дефектных изделий в партии приборов). Вычисление параметров последовательной процедуры
15. Непараметрические (свободные от распределения) критерии сдвига			
233	452	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями быстрым (грубым) критерием Кенуя
234	453	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями быстрым (грубым) ранговым критерием
235	457	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями критериями группы Манна–Уитни–Вилкоксона

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
236	460	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями критерием Фишера–Йэйтса–Терри–Гёфдинга
237	461	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями критерием Ван дер Вардена
238	463	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями медианным критерием
239	464	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями критерием Мостеллера
240	464	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями критерием Розенбаума
241	465	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями критерием Хаги
242	465	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы сдвига между двумя совокупностями E-критерием
243	468	Распределение совокупности неизвестно (более двух выборок)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Крускала–Уоллиса
244	469	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Неменьи
245	471	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Вилкоксона–Вилкокс
246	473	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями „быстрым“ критерием Кенуя
247	473	Распределение совокупности неизвестно (более двух выборок)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Фишера–Терри–Йэйтса–Гёфдинга
248	475	Распределение совокупности неизвестно (более двух выборок)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Ван дер Вардена
249	476	Распределение совокупности неизвестно (более двух выборок)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями медианным критерием
250	476	Распределение совокупности неизвестно (более двух выборок)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями против порядковой альтернативы критерием Хеттманспергера

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
251	477	Распределение совокупности неизвестно (более двух выборок)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями против порядковой альтернативы критерием Терпстры–Джонкхира
252	479	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Мостеллера
253	479	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Левиса
254	480	Распределение совокупности неизвестно (более двух выборок)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями L-критерием, основанным на U-статистиках
255	482	Распределение совокупности неизвестно (более двух выборок)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Краузе
256	483	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Пейджа
257	486	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Фридмена–Кендалла–Бэбингтона Смита
258	487	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Андерсона–Каннемана–Шэча
259	488	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Даны Квейд со взвешенными ранжировками
260	490	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Кендалла–Эренберга
261	491	Распределение совокупности неизвестно (более двух выборок равного объема)	Проверка гипотезы сдвига между несколькими (более двух) совокупностями критерием Ходжеса–Лемана–Сена
16. Непараметрические (свободные от распределения) критерии масштаба			
262	494	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы равенства параметров в двух совокупностях критерием Ансари–Бредли

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
263	496	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы равенства параметров масштаба в двух совокупностях критерием Сижела–Тьюки
264	497	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы равенства параметров масштаба в двух совокупностях критерием Кейпена
265	500	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы равенства параметров масштаба в двух совокупностях критерием Клотца
266	501	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы равенства параметров масштаба в двух совокупностях квартильным критерием
267	502	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы равенства параметров масштаба в двух совокупностях критерием Сэвиджа
268	504	Распределение совокупности неизвестно, но известно, что оно ограничено полупрямой (две выборки)	Проверка гипотезы равенства параметров масштаба в двух совокупностях критерием Муда
269	507	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы равенства параметров масштаба в двух совокупностях критерием Сукхатме
270	508	Распределение совокупности неизвестно (две парные выборки равного объема)	Проверка гипотезы равенства параметров масштаба в двух совокупностях критерием Сэндвика–Олсона
271	509	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы равенства параметров масштаба в двух совокупностях критерием Краута–Линерта, нечувствительным к сдвигу
272	511	Распределение совокупности неизвестно (две выборки)	Проверка гипотезы равенства параметров масштаба в двух совокупностях критерием Камата
273	513	Распределение совокупности неизвестно (две выборки)	Проверка комбинированной гипотезы равенства положения и масштаба в двух выборках комбинированным критерием Буша–Винда
274	515	Распределение совокупности неизвестно (более двух выборок)	Проверка гипотезы равенства параметров масштаба в нескольких (более двух) совокупностях критерием Бхапкара–Дешпанде
17. Оценка случайности эмпирических данных, выявление тренда в них (критерии тренда и случайности)			
275	518	Ряд взаимонезависимых нормально распределенных случайных величин с одинаковыми (неизвестными) дисперсиями	Проверка гипотезы случайности эмпирического ряда критерием Аббе–Линника

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
276	520	Ряд нормально распределенных случайных величин	Проверка гипотезы наличия тренда в средних и дисперсиях эмпирического ряда критерием Фостера–Стюарта
277	521	Ряд нормально распределенных случайных величин	Проверка гипотезы наличия тренда в средних и дисперсиях эмпирического ряда критерием Кокс–Стюарта
278	523	Ряд нормально распределенных случайных величин с одинаковыми средними	Проверка гипотезы наличия сдвига дисперсии в неизвестной точке критерием Хсу
279	525	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы наличия сдвига дисперсии в неизвестной точке ранговым критерием
280	529	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда критерием Вальда–Волфовитца
281	530	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда критерием Рамачандрана–Ранганатана
282	532	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда критерием Шахнесси
283	532	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда критерием Олмстеда
284	535	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда критерием числа серий знаков первых разностей
285	536	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда критерием инверсий
286	538	Ряд нормально распределенных случайных величин	Проверка гипотезы случайности эмпирического ряда критерием автокорреляции
287	539	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда критерием ранговой сериальной корреляции Вальда–Волфовитца
288	540	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда критерием Бартелса
289	541	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда критерием кумулятивной суммы
290	542	Ряд случайных величин с неизвестным законом распределения	Проверка гипотезы случайности эмпирического ряда знаково-ранговым критерием Холлина

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
18. Выявление грубых результатов наблюдений (выбросов) в выборке			
291	544	Выборка нормально распределенных случайных величин	Проверка гипотезы наличия выброса в выборке критерием Шовене
292	545	Выборка нормально распределенных случайных величин с известной дисперсией	Проверка гипотезы наличия выброса в выборке критерием Ирвина
293	546	Выборка нормально распределенных случайных величин (варианты — дисперсия известна и неизвестна)	Проверка гипотезы наличия выброса в выборке критериями Груббса
294	547	Выборка нормально распределенных случайных величин	Проверка гипотезы наличия выброса в выборке критерием наибольшего абсолютного отклонения
295	548	Выборка нормально распределенных случайных величин (дисперсия оценивается по отдельной выборке)	Проверка гипотезы наличия выброса в выборке критерием Дэвида
296	550	Выборка нормально распределенных случайных величин	Проверка гипотезы наличия выброса в выборке критерием Диксона
297	552	Выборка нормально распределенных случайных величин	Проверка гипотезы наличия выброса в выборке критериеями Хоглина–Иглевича
298	556	Выборка нормально распределенных случайных величин (число выбросов принимается заранее)	Проверка гипотезы наличия нескольких выбросов в выборке критерием Титьена–Мура
299	557	Выборка нормально распределенных случайных величин (число выбросов заранее не известно)	Проверка гипотезы наличия нескольких выбросов в выборке критерием Роснера
300	559	Выборка экспоненциально распределенных случайных величин	Проверка гипотезы наличия выбросов в выборке критерием Смоляка–Титаренко
301	561	Выборка экспоненциально распределенных случайных величин	Проверка гипотезы наличия выбросов в выборке критерием Бродского–Быцаня–Власенко
302	563	Выборка экспоненциально распределенных случайных величин	Проверка гипотезы наличия нескольких выбросов в выборке критерием Кимбера
303	565	Выборка случайных величин, имеющих распределение Вейбулла	Проверка гипотезы наличия выбросов в выборке критериями Груббса и Диксона
304	568	Выборка случайных величин из непрерывного распределения с известными параметрами	Проверка гипотезы наличия выбросов в выборке критерием Дарлинга
19. Вычисление толерантных пределов, содержащих заданную долю генеральной совокупности			
305	569	Выборка случайных величин, имеющих нормальное распределение с известными параметрами	Вычисление толерантных пределов для совокупности случайных величин

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
306	572	Выборка случайных величин, имеющих нормальное распределение с неизвестным средним и известной дисперсией	Вычисление толерантных пределов для совокупности случайных величин
307	572	Выборка случайных величин, имеющих нормальное распределение с известным средним и неизвестной дисперсией	Вычисление толерантных пределов для совокупности случайных величин
308	577	Выборка случайных величин, имеющих нормальное распределение с неизвестными параметрами	Вычисление толерантных пределов для совокупности случайных величин
309	578	Выборка случайных величин, имеющих нормальное распределение	Вычисление толерантных пределов для совокупности случайных величин с помощью выборочного размаха
310	579	Выборка случайных величин, имеющих нормальное распределение	Вычисление толерантных интервалов для выборочных дисперсий
311	582	Выборка случайных величин с неизвестным распределением	Вычисление объема выборки, в которой с заданной вероятностью будет находиться заданная доля совокупности
312	582	Выборка случайных величин с неизвестным распределением	Вычисление минимального объема выборки для одностороннего толерантного интервала
313	583	Выборка случайных величин с неизвестным распределением	Вычисление доли непрерывной совокупности, которая с заданной вероятностью будет находиться между крайними членами выборки
314	583	Выборка случайных величин, имеющих нормальное или экспоненциальное распределение	Определение границ интервала, в котором будут находиться m из k будущих наблюдений
315	588	Выборка случайных величин, имеющих экспоненциальное распределение	Вычисление прогнозных интервалов для нижней границы средней наработки по ранним моментам отказов
316	589	Выборка случайных величин, имеющих экспоненциальное распределение	Вычисление нижнего доверительного интервала для средней наработки на отказ
317	589	Выборка случайных величин, имеющих экспоненциальное распределение	Вычисление вероятностей появления заданных величин в будущих наблюдениях
318	589	Выборка случайных величин, имеющих экспоненциальное распределение	Вычисление вероятности попадания будущих наблюдений в заданный интервал
20. Дисперсионный анализ данных			
319	593	Совокупность данных, полученных при различных уровнях изучаемого фактора (распределение нормально)	Однофакторный дисперсионный анализ данных

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
320	595	Совокупность данных, полученных при различных уровнях двух факторов (распределение нормально)	Двухфакторный дисперсионный анализ данных
321	597	Совокупность данных, полученных при различных уровнях двух факторов (распределение нормально)	Двухфакторный дисперсионный анализ данных с помощью размахов
322	599	Совокупность данных, полученных при различных уровнях двух факторов (распределение неизвестно, в данных возможны пропуски)	Двухфакторный дисперсионный непараметрический анализ для неполных данных критерием Принтиса
323	602	Совокупность данных, полученных при различных уровнях двух факторов (распределение неизвестно, в данных возможны пропуски)	Двухфакторный дисперсионный непараметрический анализ для неполных данных критерием Мака–Скиллингса
324	604	Совокупность данных, полученных при различных уровнях двух факторов (распределение неизвестно, в данных возможны пропуски)	Двухфакторный дисперсионный непараметрический анализ для неполных данных критерием Лемана–Мака
21. Корреляционный анализ данных			
325	608	Две выборки нормально распределенных случайных величин	Проверка гипотезы о наличии корреляции между случайными величинами
326	610	Две выборки нормально распределенных случайных величин	Оценка корреляционного отношения (оценка линейности корреляционной связи)
327	612	Несколько (более двух) выборок нормально распределенных случайных величин	Вычисление показателей частной и множественной корреляции и оценка их значимости
328	615	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции „быстрыми“ критериями Кенуя и Кокс–Стюарта
329	617	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции знаковым критерием Нелсона
330	619	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции квадрантным критерием
331	620	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции угловым критерием Олмстеда–Тьюки
332	621	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами приближенным критерием Шахани
333	622	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами критерием Шведа–Эйзенхарта
334	623	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами критерием автокорреляции Кенуя

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
335	624	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами критерием Блума–Кифера–Розенблатта
336	625	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами с помощью рангового коэффициента корреляции Кендалла
337	628	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами с помощью рангового коэффициента корреляции Спирмена
338	629	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами критерием Гёфдинга
339	631	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами критерием Ширахатэ
340	633	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами критерием Фишера–Йэйтса
341	634	Две выборки случайных величин с неизвестным распределением	Установление значимости корреляции между случайными величинами критерием Ван дер Вардена
342	636	Несколько (больше двух) выборок случайных величин с неизвестным распределением	Установление корреляции между несколькими совокупностями случайных величин с помощью коэффициента конкордации Кендалла–Бэбингтона Смита
343	637	Несколько (больше двух) выборок случайных величин с неизвестным распределением	Установление корреляции между несколькими совокупностями случайных величин критерием Шукени–Фроли
344	638	Одна выборка случайной величины — количественная, и соответствующая ей выборка, содержащая качественные признаки	Установление значимости точечно-бисериальной корреляции (зависимости между качественным и количественным признаками)
345	640	Таблица сопряженности признаков 2×2	Установление связи признаков в таблицах 2×2
346	641	Таблица сопряженности признаков 2×2	Установление значимости связи признаков в таблицах сопряженности 2×2 с помощью коэффициента контингенции
347	642	Таблица сопряженности признаков 2×2	Установление значимости связи признаков в таблицах сопряженности 2×2 точным критерием Фишера

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
348	642	Таблица сопряженности признаков 2×2	Установление значимости связи признаков в таблицах сопряженности 2×2
349	642	Таблица сопряженности признаков 2×2	Установление значимости связи признаков в таблицах сопряженности 2×2
350	643	Таблица сопряженности признаков 2×2 (наблюдения проводятся над одним и тем же объектом)	Установление значимости связи признаков в таблицах сопряженности 2×2 модифицированным критерием знаков Мак-Нимара
351	644	Таблица сопряженности признаков 2×2	Установление значимости связи признаков в таблицах сопряженности 2×2 критерием Вулфа
352	645	Две таблицы сопряженности признаков размера 2×2 каждая	Проверка гипотезы о статистической неразличимости двух таблиц сопряженности размером 2×2 каждая критерием Ле Роя
353	646	Допустимые уровни ошибок первого и второго рода при принятии решения о наличии связи между качественными признаками	Определение объема выборки, позволяющего сравнить относительные частоты в таблицах 2×2 с заданными ошибками первого и второго рода
354	647	Таблица сопряженности признаков $r \times c$	Проверка гипотезы о наличии связи в таблице сопряженности $r \times c$
22. Регрессионный анализ данных			
355	651	Совокупность выборочных значений случайной величины, полученных при фиксированных значениях второй случайной величины	Оценка коэффициентов регрессии методом наименьших квадратов
356	651	Совокупность выборочных значений случайной величины, полученных при фиксированных значениях второй случайной величины	Вычисление коэффициентов регрессии с помощью линейных оценок
357	652	Совокупность выборочных значений случайной величины, полученных при фиксированных значениях второй случайной величины	Оценка коэффициентов регрессии методом Бартлетта–Кенуя
358	652	Совокупность выборочных значений случайной величины, полученных при фиксированных значениях второй случайной величины	Оценка коэффициентов регрессии методом Керрича
359	653	Совокупность выборочных значений случайной величины (распределение неизвестно), полученных при фиксированных значениях второй случайной величины	Оценка коэффициентов регрессии с помощью робастного критерия Брауна–Муда

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
360	654	Совокупность выборочных значений случайной величины (распределение неизвестно), полученных при фиксированных значениях второй случайной величины	Оценка регрессионного коэффициента с помощью метода Тейла
361	656	Совокупность выборочных значений случайной величины (распределение нормально), полученных при фиксированных значениях второй случайной величины	Вычисление и статистический анализ оценок коэффициентов регрессии
362	657	Совокупность выборочных значений случайной величины (распределение неизвестно), полученных при фиксированных значениях второй случайной величины	Проверка гипотезы о равенстве нулю коэффициента регрессии, оцененного с помощью метода Тейла
363	658	Совокупность выборочных значений случайной величины (распределение нормально), полученных при фиксированных значениях второй случайной величины	Проверка адекватности регрессионной модели
364	660	Совокупность выборочных значений случайной величины полученных при фиксированных значениях второй случайной величины	Проверка гипотезы о наличии корреляции регрессионных остатков критерием Дарбина–Ватсона
365	661	Совокупность выборочных значений случайной величины (распределение нормально), полученных при фиксированных значениях второй случайной величины	Проверка гипотезы о наличии выброса в регрессионной модели критерием Эктона
366	663	Совокупность выборочных значений случайной величины (распределение нормально), полученных при фиксированных значениях второй случайной величины	Проверка гипотезы о наличии выброса в регрессионной модели критерием Титьена–Мура–Бекмана
367	665	Совокупность выборочных значений случайной величины (распределение нормально), полученных при фиксированных значениях второй случайной величины	Проверка гипотезы о наличии выброса в регрессионной модели критерием Прескотта–Лунда
368	668	Совокупность выборочных значений случайной величины (распределение нормально), полученных при фиксированных значениях второй случайной величины	Построение доверительной зоны для уравнения регрессии
369	669	Совокупность выборочных значений случайной величины (распределение нормально), полученных при фиксированных значениях второй случайной величины	Оценка обращенного уравнения регрессии (с заменой переменных)

Продолжение таблицы

№	Стр.	Исходная информация	Содержание задачи
370	671	Совокупность выборочных значений случайной величины (распределение нормально), полученных при фиксированных значениях второй случайной величины	Вычисление толерантной области для линейной регрессии
371	673	Две выборки случайных величин, полученные при фиксированных значениях другой случайной величины	Проверка гипотезы о неразличимости регрессионных моделей, полученных по двум независимым выборкам
372	676	Две выборки случайных величин, порождающие регрессионные модели	Оценка доверительной области для точки пересечения двух линий регрессии
373	679	Известно стандартное отклонение изучаемой регрессионной последовательности	Определение объема испытаний для получения заданной точности оценки коэффициента регрессии
374	681	Совокупность выборочных значений случайной величины (нормальное распределение), полученных при фиксированных значениях второй случайной величины	Построение нелинейной регрессионной модели и ее статистический анализ
375	685	Совокупность выборочных значений случайной величины (нормальное распределение), полученных при фиксированных значениях второй случайной величины	Построение и статистический анализ нелинейной регрессии с помощью полиномов Чебышева
376	688	Две регрессионные модели и независимая выборка значений случайной величины	Выбор лучшей регрессионной модели критерием Вильямса–Клута
377	690	Совокупность выборочных значений случайной величины, полученных при фиксированных значениях второй случайной величины	Вычисление прогнозируемых значений случайной величины по уравнению регрессии
378	693	Двумерная выборка случайных величин	Вычисление регрессии методом наименьших модулей
379	694	Двумерная выборка случайных величин	Вычисление прямой регрессии методом последней точки
380	695	Двумерная выборка случайных величин	Вычисление регрессии методом однозначной аппроксимации
23. Вычисление параметров контрольных карт			
381	699	Выборочные значения контролируемого параметра	Вычисление границ для контрольных карт $\bar{x}$ и R-карт Шухарта
382	700	Выборочные значения контролируемого параметра	Вычисление границ для контрольных S-карт Шухарта
383	701	Выборочные значения контролируемого параметра	Вычисление контрольных границ для $\bar{x}$- и S-карт Шухарта по выборкам неравного объема

Окончание таблицы

№	Стр.	Исходная информация	Содержание задачи
384	702	Выборочные значения контролируемого параметра	Вычисление границ для контрольных карт доли дефектных изделий (P-карта Шухарта)
385	703	Выборочные значения контролируемого параметра	Вычисление границ для контрольных карт числа дефектных изделий (C-карта Шухарта)
386	703	Выборочные значения контролируемого параметра	Вычисление границ для контрольных карт индивидуальных значений и скользящего размаха (x_i- и R_i–карты Шухарта)
387	706	Выборочные значения контролируемого параметра	Вычисление параметров контрольной карты накопленных сумм для контроля среднего значения
388	708	Выборочные значения контролируемого параметра	Вычисление параметров контрольной карты накопленных сумм для контроля разброса по выборочным размахам
389	709	Выборочные значения контролируемого параметра	Вычисление параметров контрольной карты накопленных сумм для контроля разброса по выборочным дисперсиям
390	711	Выборочные значения контролируемого параметра	Вычисление параметров контрольной карты накопленных сумм для контроля доли дефектных изделий
391	712	Выборочные значения контролируемого параметра	Вычисление параметров контрольной карты накопленных сумм для контроля доли дефектных изделий, основанного на распределении Пуассона
24. Математическое планирование эксперимента, построение планов и анализ моделей			
392	718	Число изучаемых независимых параметров-факторов	Построение матрицы полного факторного эксперимента
393	719	Матрица и результаты реализации полного факторного эксперимента	Определение уравнения отклика и проведение его статистического анализа
394	726	Матрица и результаты реализации ортогонального плана второго порядка	Определение статистической модели, описывающей результаты опытов
395	731	Построение матрицы плана Хартли на кубе для пяти переменных и вычисление поверхности отклика	
396	733	Результаты последовательного экспериментирования при поиске оптимума	Определение оптимума отклика (три фактора) методом крутого восхождения Бокса–Уилсона
397	735	Результаты последовательного экспериментирования при поиске оптимума	Поиск оптимума отклика методом симплекс-планирования

Перечень математико-статистических таблиц

Продолжение таблицы

№	Стр.	Наименование	Область применения
16	112	Значения $1/d$	Оценка среднеквадратического отклонения нормальной совокупности по выборочному размаху
17	113	Значения коэффициента $f(p_1, p_2)$	Оценка среднеквадратического отклонения нормальной совокупности по шаблону
18	114	Значения $k_i \cdot (-10^4)$ для оптимальной линейной оценки σ	Оценка среднеквадратического отклонения нормальной совокупности с помощью порядковых статистик
19	114	Значения λ_i и β_i для оценки Огавы	Оценка среднеквадратического отклонения нормальной совокупности с помощью порядковых статистик
20	115	Оценки $\hat{\sigma}$ по подразмахам	Оценка среднеквадратического отклонения нормальной совокупности по размахам
21	116	Оценка σ по двум статистикам	Оценка среднеквадратического отклонения нормальной совокупности
22	118	Значения u_γ	Квантили стандартного нормального распределения для вероятностей, используемых в различных аппроксимациях
23	119	Значения квантилей распределения размаха $\omega(\gamma)$	Оценка и проверка статистических гипотез, основанных на применении размахов
24	119	Значения квантилей распределения среднего абсолютного отклонения $m(\gamma)$	Интервальная оценка среднеквадратического отклонения нормальной совокупности
25	120	Значения $\sqrt{\dfrac{n-1}{\chi_r^2}}$ для интервальной (двусторонней) оценки σ	Интервальная оценка среднеквадратического отклонения нормальной совокупности
26	122	Коэффициенты z_1 и z_2 кратчайших доверительных интервалов	Интервальная оценка среднеквадратического отклонения нормальной совокупности
27	124	Значения функции $z = f(y)$	Оценка параметров усеченного нормального распределения
28	124	Значения функции $g(z)$	Оценка параметров усеченного нормального распределения
29	125	Значения функции $z = f(h, y)$	Оценка параметров в неполностью определенных выборках
30	126	Значения функции $\psi(z)$	Оценка параметров в неполностью определенных выборках
31	127	Значения $k(h, \gamma)$	Оценка параметров нормального распределения в цензурированных выборках

Продолжение таблицы

№	Стр.	Наименование	Область применения
32	128	Коэффициенты $\chi'(\alpha, n, r)$ и $\chi''(\alpha, n, r)$	Оценка параметров нормального распределения в цензурированных выборках
33	129	Коэффициенты $t'(a, n, r)$ и $(-1) \cdot t''(\alpha, n, r)$	Оценка параметров нормального распределения в цензурированных выборках
34	130	Коэффициенты $k_\imath$ и k'_i для $n = 10$	Оценка параметров нормального распределения в цензурированных выборках с помощью порядковых статистик
35	138	Значения коэффициентов $b_\imath$ и $\varepsilon_\imath$	Оценка параметров экспоненциального распределения с помощью порядковых статистик
36	141	Значения коэффициентов $k_{\text{н}}$ и $k_{\text{в}}$	Интервальная оценка наработки на отказ
37	142	Значения коэффициентов $c'_\alpha(d)$ и $c''_\alpha(d)$ для $\alpha = 0{,}90$	Интервальная оценка интенсивности отказов
38	143	Значения коэффициентов $b'_\alpha(d)$ и $b''_\alpha(d)$ для $\alpha = 0{,}95$	Интервальная оценка интенсивности отказов
39	148	Зависимость параметра β распределения Вейбулла от коэффициента вариации $v = s/\bar{x}$	Оценка параметров распределения Вейбулла
40	149	Значения α_3, β, b и c, используемые при оценке параметров распределения Вейбулла	Оценка параметров распределения Вейбулла
41	150	Значения G, β и b' для оценки параметров распределения Вейбулла	Оценка параметров распределения Вейбулла
42	153	Значения коэффициентов $a_\imath$ и c_i	Оценка параметров распределения Вейбулла с помощью порядковых статистик
43	163	Значения коэффициентов k_1 и k_2	Оценка параметров распределения Вейбулла с помощью порядковых статистик
44	167	Значения коэффициентов c'_α и c''_α	Совместная интервальная оценка параметров распределения Вейбулла с помощью порядковых статистик
45	168	Значения коэффициентов d'_α и d''_α	Совместная интервальная оценка параметров распределения Вейбулла с помощью порядковых статистик
46	170	Коэффициенты $k'(\varepsilon, \alpha)$ и $k''(\varepsilon, \alpha)$	Интервальная оценка наработки компонента при заданной вероятности безотказной работы

Продолжение таблицы

№	Стр.	Наименование	Область применения
47	173	Значения $\eta = \frac{t(r,n)}{t(r,r)}$	Оценка сокращения продолжительности испытаний за счет цензурирования испытываемой выборки
48	196	Значения $l' = F\left(2m, 2n; \alpha^{\frac{1}{k}}\right)$ для $\alpha = 0{,}95$	Прогнозирование средней наработки до отказа в экспоненциальных выборках
49	198	Значения $\frac{t_\alpha(n)}{\sqrt{n}}$	Определение объема выборки для оценки среднего нормального распределения при неизвестной дисперсии
50	199	Значения n для распределения Вейбулла	Определение объема выборки, необходимого для оценки среднего распределения Вейбулла
51	200	Значения $a(r,\alpha)$	Планирование испытаний для оценки интенсивности отказов с заданной точностью
52	201	Значения $b(r,\alpha)$	Планирование испытаний для оценки интенсивности отказов с заданной точностью
53	215	Процентные точки статистик $\widetilde{D}_n$ и $\widetilde{D}_n^{+(-)}$	Проверка гипотезы нормальности гипотетического распределения
54	217	Квантили распределения $n\omega^2$	Проверка гипотезы нормальности гипотетического распределения
55	219	Квантили распределения χ^2	Проверка гипотезы согласия
56	220	Квантили предельного распределения статистики $n\Omega^2$	Проверка гипотезы согласия
57	222	Нижние критические точки статистики Ватсона U_n^2	Проверка гипотезы согласия гипотетического распределения с теоретическим
58	224	Критические значения $V_n'(\alpha)$ для верхних процентных точек статистики Купера	Проверка гипотезы согласия гипотетического распределения с теоретическим
59	231	Значения коэффициентов $c_\imath$ модифицированного χ^2-критерия нормальности для $k = 3 \div 15$	Проверка совпадения эмпирического распределения с нормальным
60	232	Критические значения $d_k(\alpha)$ модифицированного χ^2-критерия нормальности	Проверка совпадения эмпирического распределения с нормальным
61	233	Критические значения статистики $n\omega^2$ для проверки нормальности распределения	Проверки гипотезы нормальности распределения
62	233	Критические значения статистики Колмогорова–Смирнова, модифицированной для проверки нормальности распределения	Проверка гипотезы нормальности распределения

Продолжение таблицы

№	Стр.	Наименование	Область применения
63	235	Критические значения статистики Фроцини B_n для проверки норимальности распределения	Проверка гипотезы нормальности распределения
64	236	Коэффициенты a_{n-i+1} ($\times 10^4$) критерия Шапиро–Уилка	Проверка гипотезы нормальности распределения
65	239	Коэффициенты η, γ и ε	Аппроксимация критических точек критерия Шапиро–Уилка
66	240	Процентные точки критерия $W(\alpha)$	Проверка гипотезы нормальности распределения
67	242	Значения $K_{mn}(\alpha)$ для $\alpha = 0{,}05$	Проверка гипотезы нормальности распределения критерием Васичека
68	243	Критические значения $T_1(\alpha)$ и $T_2(\alpha)$	Проверка гипотезы нормальности распределения критерием Хегази–Грина
69	244	Критические значения $M_n(\alpha)$ критерия нормальности Али–Чёрго–Ревеса	Проверка гипотезы нормальности распределения
70	247	Критические значения $r(\alpha)$ критерия Филлибена	Проверка гипотезы нормальности распределения
71	249	Коэффициенты a_i и b_i для вычисления статистик K_1, K_2, K_3	Проверка гипотезы нормальности распределения
72	252	Критические точки критериев $K_1(\alpha)$, $K_2(\alpha)$ и $K_3(\alpha)$	Проверка гипотезы нормальности распределения
73	255	Дисперсии статистик Оя	Проверка гипотезы нормальности критерием Оя
74	258	Критические значения $d_1(\alpha)$ и $d_2(\alpha)$ критерия Гири	Проверка гипотезы нормальности критерием Гири
75	259	Критические значения $U_1(\alpha)$ и $U_2(\alpha)$ критерия Дэвида–Хартли–Пирсона	Проверка гипотезы о нормальности распределения критерием Дэвида–Хартли–Пирсона
76	260	Критические значения $T'(\alpha)$ критерия Шпигельхальтера	Проверка гипотезы нормальности распределения критерием Шпигельхальтера
77	267	Критические значения критерия Д'Агостино	Проверка гипотезы нормальности распределения
78	269	Значения коэффициентов δ и $\frac{1}{\lambda}$	Аппроксимация распределения коэффициента эксцесса нормального распределения
79	272	Критические значения a_1 и a_2 критерия Муроты–Такеучи на уровне значимости $\alpha = 0{,}10$	Проверка гипотезы нормальности распределения критерием Муроты–Такеучи
80	278	Сравнение критериев проверки нормальности распределения случайных величин	Сравнение различных критериев нормальности по мощности

Продолжение таблицы

№	Стр.	Наименование	Область применения
81	280	Критические значения критерия экспоненциальности W_E Шапиро–Уилка	Проверка гипотезы экспоненциальности распределения критерием Шапиро–Уилка
82	281	Критические значения критерия экспоненциальности W_{E_0} Шапиро–Уилка	Проверка гипотезы экспоненциальности распределения критерием Шапиро–Уилка
83	283	Критические значения статистик критериев согласия типа Колмогорова–Смирнова для проверки экспоненциальности распределения с неизвестными параметрами	Проверка гипотезы экспоненциальности распределения критериями типа Колмогорова–Смирнова
84	284	Процентные точки модифицированных критериев типа Колмогорова–Смирнова для проверки экспоненциальности распределения	Проверка гипотезы экспоненциальности распределения критериями типа Колмогорова–Смирнова
85	287	Критические значения статистики W_p^2	Проверка гипотезы экспоненциальности распределения критериями типа Крамера–фон Мизеса–Смирнова по цензурированным выборкам
86	288	Критические значения $B_n(\alpha)$ критерия экспоненциальности Фроцини	Проверка гипотезы экспоненциальности критерием Фроцини
87	289	Критические значения $K(z, m, \alpha)$ и $K(z, \tilde{m}, \alpha)$ корреляционного критерия экспоненциальности	Проверка гипотезы экспоненциальности корреляционным критерием
88	292	Критические значения $D(\alpha)$ статистики Кимбера–Мичела	Проверка гипотезы экспоненциальности критерием Кимбера–Мичела
89	295	Критические значения статистики $c = \sqrt{n}(\tilde{c} - 1)$	Проверка гипотезы экспоненциальности критерием Климко–Антла–Радемакера–Рокетта
90	296	Граничные значения $t_1(\alpha)$ и $t_2(\alpha)$ критерия Холлендера–Прошана	Проверка гипотезы экспоненциальности критерием Холлендера–Прошана
91	299	Критические значения $T_n(\alpha)$ критерия Кочара	Проверка гипотезы экспоненциальности критерием Кочара
92	300	Критические значения $c_1(\alpha)$ и $c_2(\alpha)$ критерия Эппса–Палли	Проверка гипотезы экспоненциальности критерием Эппса–Палли
93	301	Критические значения $M_n(\alpha)$ критерия Чёрго–Уэлча	Проверка гипотезы экспоненциальности критерием Чёрго–Уэлча
94	302	Значения $\mathbf{P}(L_n \geqslant k)$	Проверка гипотезы экспоненциальности критерием Бергмана
95	303	Критические значения $\omega_n(\alpha)$ статистики Шермана	Проверка гипотезы экспоненциальности критерием Шермана
96	304	Критические значения $\eta_n(\alpha)$ статистики η_n	Проверка гипотезы экспоненциальности критерием наибольшего интервала
97	305	Критические значения $h_\alpha(n)$ статистики Хартли	Проверка гипотезы экспоненциальности критерием Хартли

Продолжение таблицы

№	Стр.	Наименование	Область применения
98	309	Критические значения $\tilde{U}_\alpha$ статистики $\tilde{U}$	Проверка гипотезы экспоненциальности критерием $\bar{U}$
99	310	Критические значения $G_1(\alpha)$ и $G_2(\alpha)$ статистики Гринвуда	Проверка гипотезы экспоненциальности критерием Гринвуда
100	312	Коэффициенты a_i K–критерия согласия Манн–Фертига–Шуера для распределения Вейбулла	Проверка гипотезы согласия эмпирического распределения с распределением Вейбулла
101	314	Критические значения $K_\alpha(r, n)$ критерия Манн–Фертига–Шуера	Проверка гипотезы согласия эмпирического распределения с распределением Вейбулла
102	318	Критические значения $W_1(\alpha)$ и $W_2(\alpha)$ критерия экспоненциальности Лоулесса	Проверка гипотезы экспоненциальности критерием Лоулесса
103	321	Критические значения $M_1(\alpha)$ и $M_2(\alpha)$ статистики Морана	Проверка гипотезы равномерности распределения критерием Морана
104	323	Критические значения $W_1(\alpha)$ и $W_2(\alpha)$ критерия равномерности Ченга–Спиринга	Проверка гипотезы равномерности распределения критерием Ченга–Спиринга
105	324	Критические значения $J(\alpha)$ критерия Саркади–Косика	Проверка гипотезы равномерности распределения критерием Саркади–Косика
106	325	Критические значения $H_\alpha(m, n)$ статистики Дудевича–ван дер Мюлена	Проверка гипотезы равномерности критерием Дудевича–ван дер Мюлена
107	327	Критические значения $T_1(\alpha)$, $T_1^*(\alpha)$, $T_2(\alpha)$ и $T_2^*(\alpha)$ критерия равномерности Хегази–Грина	Проверка гипотезы равномерности распределения критерием Хегази–Грина
108	329	Критические значения $M_1(\alpha)$, $M_2(\alpha)$ статистики Янга	Проверка гипотезы равномерности распределения критерием Янга
109	330	Критические значения $\tilde{D}^+$, $\tilde{D}^-$, $\tilde{D}$, $\tilde{V}$ критериев равномерности	Проверка гипотезы равномерности критериями Колмогорова–Смирнова
110	331	Критические значения $B_n(\alpha)$ критерия Фроцини для проверки равномерности распределения	Проверка гипотезы равномерности распределения критерием Фроцини
111	332	Критические значения $Q(\alpha)$ статистики Гринвуда–Кэсенберри–Миллера для проверки равномерности распределения	Проверка гипотезы равномерности распределения критерием Гринвуда–Кэсенберри–Миллера
112	334	Критические значения $N_K(\alpha)$ критерия равномерности Неймана–Бартона	Проверка гипотезы равномерности распределения критерием Неймана–Бартона
113	338	Критические значения $K(\alpha)$ критерия знаков	Проверка гипотезы симметрии эмпирического распределения критерием знаков

Продолжение таблицы

№	Стр.	Наименование	Область применения
114	339	Критические значения $T^+(\alpha)$ одновыборочного критерия Вилкоксона	Проверка гипотезы симметрии эмпирического распределения критерием Вилкоксона
115	346	Критические значения B-критерия Бооса	Проверка гипотезы симметрии эмпирического распределения критерием Бооса
116	358	Оценки параметров γ и η распределения S_U Джонсона	Аппроксимация эмпирического распределения кривой из семейства S_U Джонсона
117	370	Значения квантилей $y_\alpha = (x_\alpha - \bar{x})/s$ нормированного распределения Пирсона	Нахождение квантилей эмпирического распределения с помощью кривых распределения Пирсона
118	391	Таблица квантилей распределения Стьюдента t_γ	Проверка статистических гипотез
119	393	Критические значения модифицированного критерия Стьюдента для сравнения средних по двум выборкам равного объема	Сравнение средних значений двух нормально распределенных выборок модифицированным критерием Стьюдента
120	395	Критические значения $N_0(r,n)$ для доверительной вероятности $\alpha = 0{,}95$	Сравнение средних значений двух нормально распределенных выборок двухступенчатым критерием Волфа
121	396	Значения r и N_0 при $n \geqslant 20$ для двухступенчатого двухвыборочного критерия сравнения средних при доверительной вероятности $\alpha = 0{,}95$	Сравнение средних значений двух нормально распределенных выборок двухступенчатым критерием Волфа
122	398	Критические значения $Q_\alpha(k,n)$ модифицированного критерия Стьюдента	Сравнение средних нескольких (> 2) нормально распределенных выборок равного объема
123	400	Критические значения $q_\alpha(n,f)$ „стьюдентизированного“ размаха	Сравнение средних нескольких (> 2) нормально распределенных выборок равного объема
124	403	Значения $F'_\alpha(1; kn-2)$ для доверительной вероятности $\alpha = 0{,}95$	Сравнение средних нескольких (> 2) нормально распределенных выборок равного объема критерием Полсона
125	404	Критические значения T_α критерия Тьюки	Сравнение средних нескольких (> 2) нормально распределенных выборок равного объема методом прямого сравнения (критерий Тьюки)
126	406	Критические значения $m_{\alpha,k^*,\nu}$ „стьюдентизированного“ максимума	Сравнение средних нескольких (> 2) нормально распределенных выборок равного объема критерием „стьюдентизированного“ максимума
127	409	Множители $q(k^*, f, \alpha)$ для критерия Стьюдента–Ньюмена–Кейлса при доверительной вероятности $\alpha = 0{,}95$	Сравнение средних нескольких (> 2) нормально распределенных выборок равного объема критерием Стьюдента–Ньюмена–Кейлса

Продолжение таблицы

№	Стр.	Наименование	Область применения
128	410	Множители $q(k^*, f, \alpha')$ для критерия Дункана при доверительной вероятности $\alpha = 0{,}95$	Сравнение средних нескольких (> 2) нормально распределенных выборок равного объема критерием Дункана
129	411	Критические значения $K(n, k, \alpha)$ для критерия Линка–Уоллеса при доверительной вероятности $\alpha = 0{,}95$	Сравнение средних нескольких (> 2) нормально распределенных выборок равного объема критерием Линка–Уоллеса
130	414	Критические значения $F^*(n, m)$ отношения размахов	Сравнение дисперсий двух нормальных выборок критерием отношения размахов
131	416	Критические значения $K_1(\alpha)$ и $K_2(\alpha)$ статистики Аризоно–Охты для доверительной вероятности $\alpha = 0{,}95$	Сравнение дисперсий двух нормально распределенных выборок критерием Аризоно–Охты
132	419	Критические значения $g_\alpha(k, n)$ статистики Кохрана для доверительной вероятности $\alpha = 0{,}95$	Сравнение дисперсий нескольких (> 2) нормально распределенных выборок равного объема критерием Кохрана
133	420	Критические значения H_α критерия Неймана–Пирсона для доверительной вероятности $\alpha = 0{,}95$	Сравнение дисперсий нескольких (> 2) нормально распределенных выборок равного объема критерием Неймана–Пирсона
134	420	Критические значения H_α^* критерия Неймана–Пирсона для доверительной вероятности $\alpha = 0{,}95$	Сравнение дисперсий нескольких (> 2) нормально распределенных выборок равного объема критерием Неймана–Пирсона
135	421	Критические значения $c_\alpha(n, k)$ критерия Блисса–Кохрана–Тьюки для доверительной вероятности $\alpha = 0{,}95$	Сравнение дисперсий нескольких (> 2) нормально распределенных выборок равного объема критерием Блисса–Кохрана–Тьюки
136	422	Критические значения $h_\alpha(n, k)$ критерия Хартли для доверительной вероятности $\alpha = 0{,}95$	Сравнение дисперсий нескольких (> 2) нормально распределенных выборок равного объема критерием Хартли
137	422	Критические значения $K_\alpha(n, k)$ критерия Кэдуэлла–Лесли–Брауна для доверительной вероятности $\alpha = 0{,}95$	Сравнение дисперсий нескольких (> 2) нормально распределенных выборок равного объема критерием Кэдуэлла–Лесли–Брауна
138	427	Значения коэффициентов $K_\alpha(c)$	Сравнение выборочной интенсивности отказов (параметр экспоненциального распределения) с заданным значением
139	430	Критические значения D_α критерия Дэвида	Сравнение параметров нескольких (> 2) экспоненциальных распределений

Продолжение таблицы

№	Стр.	Наименование	Область применения
140	431	Критические значения $L_\alpha(n,k)$ критерия отношения правдоподобия для доверительной вероятности $\alpha = 0{,}95$	Сравнение параметров нескольких (> 2) экспоненциальных распределений
141	432	Критические значения $c_\alpha(k,n)$ статистики Чена	Сравнение гарантийных наработок нескольких экспоненциально распределенных выборок критерием Чена
142	440	Функция распределения $W_c(x)$ Вальда	Последовательные методы проверки гипотез о значениях параметров распределений критерием Вальда
143	455	Критические значения $U_1(\alpha)$ и $U_2(\alpha)$ критерия Манна-Уитни	Сравнение параметров сдвига двух совокупностей критерием Манна-Уитни
144	457	Критические значения статистики T знакового рангового критерия Вилкоксона	Сравнение параметров сдвига двух совокупностей с неизвестным распределением критерием Вилкоксона
145	459	Критические значения статистики Фишера–Йэйтса–Терри–Гёфдинга	Сравнение параметров сдвига совокупностей с неизвестным распределением критерием Фишера–Йэйтса–Терри–Гёфдинга
146	461	Критические значения X_α статистики Ван дер Вардена	Сравнение параметров сдвига двух совокупностей с неизвестным распределением критерием Ван дер Вардена
147	462	Вспомогательные величины K критерия Ван дер Вардена	Вычисление статистики критерия Ван дер Вардена (см. выше)
148	465	Критические значения T_α критерия Хаги	Сравнение параметров сдвига двух совокупностей с неизвестным распределением критерием Хаги
149	466	Критические значения критерия Крускала–Уоллиса	Сравнение параметров нескольких (> 2) совокупностей с неизвестным распределением критерием Крускала–Уоллеса
150	470	Критические значения $D_{l,m}(\alpha)$ критерия Неменьи	Сравнение параметров нескольких (> 2) совокупностей с неизвестным распределением критерием Неменьи
151	472	Критические значения $D_{\nu,\varepsilon}(\alpha)$ критерия Вилкоксона–Вилкокс	Сравнение параметров нескольких (> 2) совокупностей с неизвестным распределением критерием Вилкоксона–Вилкокс
152	473	Критические значения $n(\alpha)$ критерия Кенуя	Сравнение параметров нескольких (> 2) совокупностей с неизвестным распределением критерием Кенуя
153	478	Критические значения S_α критерия Терпстры–Джонкхира для $k = 3$	Сравнение параметров нескольких (> 2) совокупностей с неизвестным распределением критерием Терпстры–Джонкхира

Продолжение таблицы

№	Стр.	Наименование	Область применения
154	478	Критические значения S_α критерия Терпстры–Джонкхира для выборок равного объема	Сравнение нескольких (> 2) совокупностей по выборкам равного объема критерием Терпстры–Джонкхира
155	480	Критические значения L_α для $k = 3$ и доверительной вероятности $\alpha = 0{,}95$	Сравнение параметров нескольких (> 2) совокупностей с неизвестным распределением L-критерием
156	483	Критические значения $L_\alpha(k, n)$ критерия Пейджа для доверительной вероятности $\alpha = 0{,}95$	Сравнение параметров нескольких (> 2) совокупностей с неизвестным распределением по выборкам равного объема критерием Пейджа
157	485	Критические значения $S_\alpha(n, k)$ критерия Фридмана–Кендалла–Бэбингтона Смита	Сравнение параметров нескольких (> 2) совокупностей с неизвестным распределением по выборкам равного объема критерием Фридмана–Кендалла–Бэбингтона Смита
158	489	Критические значения $K_\alpha(k, n)$ критерия Кендалла–Эренберга для доверительной вероятности $\alpha = 0{,}95$	Сравнение параметров нескольких (> 2) совокупностей с неизвестным распределением по выборкам равного объема критерием Кендалла–Эренберга
159	494	Критические значения $S_1(\alpha)$ и $S_2(\alpha)$ статистики Ансари–Бредли	Сравнение параметров масштаба двух совокупностей с неизвестным распределением критерием Ансари–Бредли
160	497	Метки $a_N(i)$ критерия Кейпена	Сравнение параметров масштаба двух совокупностей с неизвестным распределением критерием Кейпена
161	498	Критические значения $K_1(\alpha)$ и $K_2(\alpha)$ статистики Кейпена	Сравнение параметров масштаба двух совокупностей с неизвестным распределением критерием Кейпена
162	499	Метки $U^2_{\frac{i}{N}}$ критерия Клотца	Сравнение параметров масштаба двух совокупностей с неизвестным распределением критерием Клотца
163	500	Критические значения $L_1(\alpha)$ и $L_2(\alpha)$ критерия Клотца	Сравнение параметров масштаба двух совокупностей с неизвестным распределением критерием Клотца
164	503	Критические значения $C_1(\alpha)$ и $C_2(\alpha)$ критерия Сэвиджа	Сравнение параметров масштаба двух совокупностей с неизвестным распределением критерием Сэвиджа
165	505	Критические значения $M_1(\alpha)$ и $M_2(\alpha)$ статистики Муда	Сравнение параметров масштаба двух совокупностей с неизвестным распределением критерием Муда
166	510	Критические значения $D_1(\alpha)$ и $D_2(\alpha)$ масштабного критерия Камата	Сравнение параметров масштаба двух совокупностей с неизвестным распределением критерием Камата

Продолжение таблицы

№	Стр.	Наименование	Область применения
167	511	Значения корректирующих коэффициентов $K_1(0{,}95)$ и $K_2(0{,}95)$ критерия Камата	Сравнение параметров масштаба двух совокупностей с неизвестным распределением критерием Камата
168	512	Критические значения $W_{mn}(\alpha)$ статистики Буша–Винда для $m = n$	Сравнение параметров положения и масштаба двух совокупностей с неизвестным распределением комбинированным критерием Буша–Винда
169	515	Критические значения L_α и D_α для $k = 3$	Сравнение параметров положения нескольких совокупностей с неизвестным распределением
170	518	Критические значения q_α критерия Аббе–Линника	Проверка гипотезы случайности выборочного ряда
171	519	Постоянные f и l критерия Фостера–Стюарта	Проверка гипотезы случайности выборочного ряда
172	522	Критические значения (H_α) H^*-критерия Хсу	Проверка гипотезы сдвига дисперсии выборочного ряда в неизвестной точке
173	523	Критические значения (G_α) G-критерия Хсу	Проверка гипотезы сдвига дисперсии выборочного ряда в неизвестной точке
174	525	Критические значения рангового критерия обнаружения сдвига дисперсии	Проверка гипотезы сдвига дисперсии выборочного ряда в неизвестной точке
175	527	Критические значения $N_1(\alpha)$ и $N_2(\alpha)$ сериального критерия Вальда–Волфовитца	Проверка гипотезы случайности выборочного ряда
176	530	Критические значения $R(\alpha)$ критерия Рамачандрана–Ранганатана	Проверка гипотезы случайности выборочного ряда
177	531	Критические значения $N_\alpha(n_\imath, k)$ сериального множественного критерия Шахнесси для $k = 2 \div 6$	Проверка гипотезы случайности выборочного ряда
178	533	Критические значения $n_\alpha(l_\imath)$ критерия случайности Олмстеда	Проверка гипотезы случайности выборочного ряда
179	534	Критические значения $R_1(\alpha)$ и $R_2(\alpha)$ критерия числа серий знаков первых разностей	Проверка гипотезы случайности выборочного ряда
180	536	Критические значения $I_1(\alpha)$ и $I_2(\alpha)$ критерия инверсий	Проверка гипотезы случайности выборочного ряда
181	537	Критические значения $r'_{1,n}(\alpha)$ и $r''_{1,n}(\alpha)$ критерия автокорреляции	Проверка гипотезы случайности выборочного ряда
182	540	Коэффициенты аппроксимации критического значения критерия Бартелса	Проверка гипотезы случайности выборочного ряда
183	541	Критические значения $R_1(\alpha)$ и $R_2(\alpha)$ критерия кумулятивной суммы	Проверка гипотезы случайности выборочного ряда

Продолжение таблицы

№	Стр.	Наименование	Область применения
184	542	Значения k для знаково-рангового критерия случайности Холлина	Проверка гипотезы случайности выборочного ряда
185	543	Критические значения r_α знаково-рангового критерия случайности Холлина	Проверка гипотезы случайности выборочного ряда
186	544	Критические значения K^* критерия Шовене	Проверка гипотезы наличия выброса в выборке, имеющей нормальное распределение
187	545	Критические значения $\tau(\alpha)$ критерия Ирвина	Проверка гипотезы наличия выброса в выборке, имеющей нормальное распределение с известной дисперсией
188	546	Критические значения $\tau_1(\alpha)$, $\tau_2(\alpha)$ и $\tau_3(\alpha)$ статистик Груббса	Проверка гипотезы наличия выброса в выборке, имеющей нормальное распределение
189	547	Критические значения $\tau_4(\alpha)$ критерия наибольшего абсолютного отклонения	Проверка гипотезы наличия выброса в выборке, имеющей нормальное распределение
190	549	Критические значения $\tau_5(\alpha)$ статистики Дэвида	Проверка гипотезы наличия выброса в выборке, имеющей нормальное распределение
191	551	Критические значения $r_{10}(\alpha)$, $r_{11}(\alpha)$, $r_{12}(\alpha)$, $r_{20}(\alpha)$, $r_{21}(\alpha)$ и $r_{22}(\alpha)$ статистик Диксона	Проверка гипотезы наличия выброса в выборке, имеющей нормальное распределение
192	553	Значения коэффициента k критерия Хоглина–Иглевича	Проверка гипотезы наличия выброса в выборке, имеющей нормальное распределение
193	554	Критические значения $L_k(\alpha)$ и $L_k^*(\alpha)$ критерия Титьена-Мура	Проверка гипотезы наличия нескольких выбросов в выборке, имеющей нормальное распределение
194	555	Критические значения $E_k(\alpha)$ критерия Титьена–Мура	Проверка гипотезы наличия нескольких выбросов в выборке, имеющей нормальное распределение
195	558	Критические значения $\tau_{1\imath}^*$ критерия Роснера	Проверка гипотезы наличия нескольких выбросов (количество выбросов априори неизвестно) в выборке, имеющей нормальное распределение
196	560	Критические значения $C_k(\alpha)$ критерия Смоляка–Титаренко	Проверка гипотезы наличия выбросов в выборке, имеющей экспоненциальное распределение
197	562	Критические значения $S_\jmath(\alpha)$ критерия Кимбера для выделения j верхних выбросов	Проверка гипотезы наличия нескольких выбросов в выборке, имеющей экспоненциальное распределение

Продолжение таблицы

№	Стр.	Наименование	Область применения
198	564	Критические значения $S_j^*(\alpha)$ критерия Кимбера для выделения j нижних выбросов	Проверка гипотезы наличия нескольких выбросов в выборке, имеющей экспоненциальное распределение
199	566	Критические значения $G(\alpha)$ и $R_i(\alpha)$ критериев типа Груббса и Диксона для выявления k верхних выбросов в выборке, имеющей распределение Вейбулла	Проверка гипотезы наличия нескольких выбросов в выборке, имеющей распределение Вейбулла
200	567	Критические значения $G^*(\alpha)$ и $R_i^*(\alpha)$ критериев типа Груббса и Диксона для выявления k нижних выбросов в выборке, имеющей распределение Вейбулла	Проверка гипотезы наличия нескольких выбросов в выборке, имеющей распределение Вейбулла
201	570	Множители $a(n,\alpha,\beta)$ для построения односторонних толерантных пределов в случае нормального распределения (μ неизвестно, σ^2 известно)	Вычисление толерантных пределов нормально распределенной случайной величины (среднее неизвестно, дисперсия известна)
202	571	Множители $a^*(n,\alpha,\beta)$ для построения двусторонних толерантных интервалов в случае нормального распределения (μ неизвестно, σ^2 известно)	Вычисление толерантных пределов нормально распределенной случайной величины (среднее неизвестно, дисперсия известна)
203	573	Множители $b(n,\alpha,\beta)$ для построения односторонних толерантных пределов в случае нормального распределения (μ известно, σ^2 неизвестно)	Вычисление толерантных пределов нормально распределенной случайной величины (среднее известно, дисперсия неизвестна)
204	574	Множители $b^*(n,\alpha,\beta)$ для построения двусторонних толерантных интервалов в случае нормального распределения (μ известно, σ^2 неизвестно)	Вычисление толерантных пределов нормально распределенной случайной величины (среднее известно, дисперсия неизвестна)
205	575	Множители $k(n,\alpha,\beta)$ для построения односторонних толерантных пределов в случае нормального распределения (μ и σ^2 неизвестны)	Вычисление толерантных пределов нормально распределенной случайной величины (среднее и дисперсия неизвестны)
206	576	Множители $k^*(n,\alpha,\beta)$ для построения двусторонних толерантных интервалов в случае нормального расределения (μ и σ^2 неизвестны)	Вычисление толерантных пределов нормально распределенной случайной величины (среднее и дисперсия неизвестны)
207	577	Множители $d(1,n,\alpha,\beta)$ для определения двустороннего толерантного интервала с использованием размахов (при $m=1$, без разбиения на подвыборки)	Вычисление толерантных интервалов нормально распределенной случайной величины с помощью размахов

Продолжение таблицы

№	Стр.	Наименование	Область применения
208	578	Множители $d(m, n', \alpha, \beta)$ для определения двустороннего толерантного интервала с использованием размахов	Вычисление толерантных интервалов нормально распределенной случайной величины с помощью размахов
209	580	Минимальные объемы выборок для построения непараметрических (двустороннего, $n_{\min}$ и одностороннего, $n^*_{\min}$) толерантных интервалов, основанных на наибольших и наименьших выборочных значениях	Вычисление толерантных интервалов для случайной величины с неизвестным законом распределения вероятностей
210	581	Наименьшая доля β совокупности, заключенная внутри непараметрического двустороннего толерантного интервала (между $x_{\min}$ и $x_{\max}$ в выборке объема n)	Вычисление доли совокупности с неизвестным законом распределения вероятностей, заключенной между крайними значениями выборки
211	582	Наименьшая доля β совокупности, заключенная внутри непараметрического одностороннего толерантного интервала (больше $x_{\min}$ или меньше $x_{\max}$ в выборке объема n)	Вычисление доли совокупности с неизвестным законом распределения вероятностей, заключенной вне крайних значений выборки
212	584	Коэффициенты $r(m, n, k, \alpha)$, для которых вероятность того, что в m или более из будущих k наблюдений нормального распределения величина x будет превышать $\bar{x} - r\,(m, n, k, \alpha)\, s$, равна α	Прогнозирование значений будущих наблюдений нормально распределенной совокупности по результатам наблюдения над выборкой
213	585	Коэффициенты $b(m, n, k, \alpha)$, для которых вероятность того, что в m или более из будущих k наблюдений экспоненциального распределения величина x будет превышать $b\,(m, n, k, \alpha) \sum\limits_{i=1}^{n} x_i$, равна α	Прогнозирование значений будущих наблюдений экспоненциально распределенной совокупности по результатам наблюдения над выборкой
214	588	Значения $c(k, n, m, \alpha)$ для прогнозирования при $\alpha = 0{,}95$ нижней границы средней наработки в k выборках объема n по m ранним моментам отказов	Прогнозирование результатов испытаний изделий на надежность
215	597	Масштабный коэффициент c и эффективное число степеней свободы f для двухфакторного дисперсионного анализа	Двухфакторный дисперсионный анализ с использованием размахов
216	602	Критические значения $T_\alpha(m, k, n_{ji})$ статистики Мака–Скиллингса	Двухфакторный дисперсионный анализ при неизвестном распределении данных и при наличии пропущенных наблюдений

Продолжение таблицы

№	Стр.	Наименование	Область применения
217	608	Критические значения r_α выборочного коэффициента корреляции для $\rho = 0$	Проверка гипотезы о наличии корреляции между случайными величинами
218	612	Критические значения $r_{1,23\ldots k}$ коэффициента множественной корреляции	Проверка гипотезы о наличии множественной корреляции между случайными величинами
219	614	Критические значения n_α критерия Кенуя	Оценка корреляции между переменными с помощью порядковых статистик
220	615	Критические значения T_α критерия Кокс–Стюарта	Оценка корреляции между переменными с помощью порядковых статистик
221	616	Критические значения S_α знакового критерия корреляции Нелсона	Оценка корреляции между случайными величинами
222	618	Критические значения $S_1(\alpha)$ и $S_2(\alpha)$ квадрантного критерия корреляции	Оценка корреляции между случайными величинами
223	620	Критические значения Q_α угловой статистики Олмстеда–Тьюки	Оценка корреляции между случайными величинами
224	622	Критические значения m_α критерия Шведа–Эйзенхарта	Оценка корреляции между случайными величинами
225	623	Критические значения N_α критерия автокорреляции Кенуя	Оценка корреляции между случайными величинами
226	629	Критические значения D_α критерия независимости Гёфдинга	Проверка гипотезы о независимости случайных величин
227	631	Критические значения S_α критерия Ширахатэ	Проверка гипотезы о наличии корреляции
228	635	Критические значения $S_W(\alpha)$ для коэффициента конкордации W	Проверка гипотезы о наличии связей (согласованности) для нескольких последовательностей
229	646	Значения коэффициента K	Выбор числа наблюдений для анализа таблиц сопряженности 2×2
230	649	Линеаризующие функциональные преобразования $(y^* = a^* + b^*x^*)$	Выбор типа кривой регрессии
231	659	Критические значения статистики Дарбина–Ватсона	Проверка гипотезы о наличии корреляции регрессионных остатков
232	661	Критические значения V_α критерия выбросов Эктона	Проверка гипотезы о наличии выброса в регрессионной моделили
233	662	Критические значения R_α критерия выбросов в регрессии Титьена–Мура–Бекмана	Проверка гипотезы о наличии выброса в регрессионной модели
234	664	Критические значения R^*_α критерия выбросов в регрессии Прескотта–Лунда	Проверка гипотезы о наличии выброса в регрессионной модели

Окончание таблицы

№	Стр.	Наименование	Область применения
235	666	Коэффициенты $u(\gamma, \alpha, \lambda)$ для построения доверительной области линейной регрессии	Построение доверительной области для линейной регрессии
236	679	Значения коэффициента $k(\alpha, n)$	Определение объема испытаний для получения заданной точности оценки регрессии
237	684	Полиномы Чебышева $P_j(i)$ и значения $P^* = \sum_{i=1}^{n} P_k^2(i)$	Вычисление и анализ нелинейных регрессионных моделей
238	698	Коэффициенты для вычисления границ регулирования контрольных карт Шухарта	Построение контрольных карт Шухарта для регулирования технологических процессов
239	700	Значения коэффициентов A_i для $\bar{x}$-карт Шухарта при неравных объемах контролируемых выборок	Прстроение контрольных карт Шухарта для регулирования технологических процессов по выборкам неравного объема
240	708	Коэффициенты c и ν для ККНС выборочного размаха	Построение ККНС для контроля разброса случайной величины по размаху
241	712	Относительная эффективность различных контрольных карт	Выбор типа контрольной карты для контроля технологического процесса
242	717	Матрица полного факторного эксперимента (ПФЭ) 2^3	Математическое планирование экстремального эксперимента
243	721	Дробные реплики 2^{3-1}	Математическое планирование экстремального эксперимента
244	725	Параметры ортогональных композиционных планов	Математическое планирование экстремальных экспериментов
245	728	Параметры униформ-ротатабельных планов	Математическое планирование экстремального эксперимента
246	729	Характеристики D-оптимального плана	Математическое планирование экстремального эксперимента
247	731	Параметры планов Хартли	Математическое планирование экстремального эксперимента

Предметный указатель

Именной указатель

Научное издание

КОБЗАРЬ Александр Иванович

ПРИКЛАДНАЯ МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. ДЛЯ ИНЖЕНЕРОВ И НАУЧНЫХ РАБОТНИКОВ

Редактор *В.С. Аролович*
Оригинал-макет: *А.А. Пярнпуу*
Оформление переплета: *А.Ю. Алехина, Л.А. Гусева*

Подписано в печать 05.05.06. Формат 70×100/16. Бумага офсетная.
Печать офсетная. Усл. печ. л. 65,79. Уч.-изд. л. 72,4. Тираж 1500 экз. Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117997, Москва, ул. Профсоюзная, 90
E-mail: fizmat@maik.ru, fmlsale@maik.ru;
http://www.fml.ru

Отпечатано с готовых диапозитивов
в ОАО «Московская типография № 6»
115088, г. Москва, Ж-88, ул. Южнопортовая, 24

ISBN 5-9221-0707-0

9 785922 107075